

Don't boycott Israel's scientists

A petition urging European institutions to initiate moratoria against scientific collaboration with Israel has sparked further petitions and counter-petitions. Such boycotts are misguided and should be opposed in favour of constructive initiatives.

Science is less political than other issues, and is a bridge for peace. That is what Leah Boehm, then chief scientist at Israel's science ministry, enthusiastically told *Nature* in 1995. Then, Israeli and Palestinian researchers were optimistic that the peace process would cause funds to flow to joint Arab-Israeli projects from the international community, reinforcing peace by contributing to dialogue, and boosting research in the region to better manage water and other scarce local resources, protect the environment and fight disease.

Subsequent events have left these noble aspirations in tatters. If the world's scientific community can do anything to encourage Middle-East peace they should jump at it. Unfortunately, initiatives taken by some academics are running counter to these aims. Steven and Hilary Rose, from Britain's Open University, instigated a petition in a UK newspaper, *The Guardian*, calling for a European boycott of research and cultural links with Israel until the latter "abides by UN resolutions and opens serious peace negotiations with the Palestinians". Some 300 academics have already signed this petition, while broader boycott initiatives have since sprung up in France and several other countries.

Ariel Sharon has already defied the United Nations and most of the international community; it is naive to think he is going to lose a moment's sleep over his country's researchers being turned into pariahs. More importantly, as is summed up well in many of the responses and counter-petitions that have been launched (see, for example, <http://euroisrael.huji.ac.il/news.html> and the Correspondence on page 15 of this issue), the concept of a research boycott restricts channels that are better kept open. Ironically, the majority of Israeli

scientists are on the political left and support the peace process. Why should they be punished for the excesses of their political leaders?

The boycott is also partisan. Scientists have as much right to express their diverse views on the Israeli–Palestinian conflict as other citizens, but the 'logic' of politically motivated boycotts is too readily extensible to the point of absurdity — should we also boycott Palestinian researchers because the Palestinian Authority has not done enough to prevent suicide bombers?

Many Israeli scientists are open to collaboration with Palestinian and other Arab scientists. Take the joint graduate programme between Tel Aviv University and Bethlehem University involving geneticist Mary-Claire King at the University of Washington. This project is looking at the genetic origins of deafness, and the Middle East is ideal for such research, as rich pedigrees are available because marrying close relatives is unusually common. Moreover, Bethlehem researchers have spent time in King's laboratory, bringing back new skills to Palestine. King, who knows the realities on the ground, opposes the boycott, arguing that the three-way collaboration is an example of a successful effort to maintain such constructive contact.

This is just one example of partnerships involving Israel, its neighbours and scientists in Europe or the United States. Israel is a research powerhouse that, given an eventual improvement of relations with its neighbours, could rejuvenate science and development in the region through collaboration and training. Rather than signing boycotts, which will achieve nothing, researchers worldwide can help the peace process concretely by actively initiating more of such three-way collaborations — and encouraging their institutions to do the same. ■

Selling science to the young

More needs to be done to tell people about the rewards of careers outside science as well as inside it.

There is a problem with the younger generation. Across the world, young people in schools are choosing not to experiment with dangerous chemicals, many are failing to gaze into space, and some are even losing interest in sexual reproduction. Clearly something has to be done.

Falling interest in science — the physical sciences in particular — among teenagers is of global concern. Recent reports have highlighted a collapse in the number of students choosing to study chemistry and physics at university: UK chemistry students are down by 25% and physics students in France have fallen by 46% over recent years. Germany, the United States and Japan face similar problems.

Some of the reasons for this are well rehearsed. Science is universally acknowledged to be a relatively difficult subject, and at school level it is crammed with facts, leaving little room for creativity or debate. Some are already trying to strip down science curricula to address this.

Other reasons for the malaise given in individual countries range from the decline in experimental work in science lessons to the decision to bundle chemistry, physics and biology together to be taught as 'science', as well as low pay and poor-quality teaching.

The proposed solutions are equally familiar and often revolve

around better 'engaging' young people with science. Typical schemes see postgraduate students visit local schools and participate in science lessons. These are noble attempts, but the lure of a career and ready cash should not be underestimated — increasing numbers of students choose their degrees, and hence their school subjects, with these in mind. How many students chose to become well-paid accountants because they were 'engaged' with that subject?

Nature would be the first to point out that few researchers working in cash-strapped universities are in it for the money. Nevertheless, many graduates do turn their science degrees into decent financial as well as spiritual reward. Are students aware of this? A recent review of the supply of scientists and engineers in Britain suggested not, pointing out that few pupils make a link between what they view as vocational scientific subjects and the possibility of alternative future careers.

This perception needs to be reversed. Scientific societies should do more to highlight those who have used a scientific training as a springboard to lucrative careers outside their discipline. Those giving careers advice in schools must ensure that students see the well-paid opportunities for those with science backgrounds, not just in research but in finance, business — even, dare we say it, the media. ■





Flight delays
Large-scale study of
air travel gets off to
turbulent start
p4



Hawaii says no
Telescope expansion
plans land NASA in
court
p5



Silent satellite
Launch delay leaves
sky surveys in
jeopardy
p6



Groundwork
Afghanistan set to
benefit from import
of seeds
p7

Venter lays foundations for fresh career as ethical philanthropist

Declan Butler

Bioethics and ecology are the new watchwords for the man behind the privately funded sequence of the human genome.

Craig Venter, who stepped down in January after three years as president of Celera Genomics, the biotechnology company in Rockville, Maryland, says he plans to devote his immediate future to running a new foundation also based in Rockville.

He has announced the creation of the not-for-profit J. Craig Venter Science Foundation. This will support The Institute for Genomic Research (TIGR) — the existing research operation run by Venter's wife, Claire Fraser — and two new institutes, the TIGR Center for the Advancement of Genomics, a bioethics think-tank, and the Institute for Biological Energy Alternatives.

"I'm delighted to be out of the business world, and I'm much happier being back in the basic science not-for-profit world where I'm free to speak out without worrying about



Profit-shy: Craig Venter wants to concentrate on renewable energy and the ethics of genetics.

offending shareholders or those with vested interests," Venter says. The endowment

reportedly kicks off with \$100 million of Venter's own money.

The bioethics think-tank will assess issues such as genetic privacy, discrimination, the genetics of ethnicity and stem cells, and will have a public-outreach and educational component. Venter claims that the availability of the human genome sequence is fuelling a resurgence in the view that genes can determine all aspects of human behaviour, and says that the think-tank will try to untangle what is deterministic and what is not.

The Institute for Biological Energy Alternatives will research clean energy sources and the sequestration of carbon dioxide, Venter says. It will use microbial genomics to engineer metabolic pathways that could generate microbes with enhanced capacities for producing hydrogen and fixing carbon dioxide.

In his spare time Venter says he plans to "spend some time examining my own genetic code". He has revealed that he was one of the six supposedly anonymous donors used to generate Celera's genome sequence. The published sequence is a composite, but Venter says he has access from the Celera database to his own genome.

Not everyone is amused by the revelation. Arthur Caplan, a bioethicist at the University of Pennsylvania, and member of Celera's scientific advisory board, says it is "not unethical, but a disappointment to me. My hope was that the genome would be seen as something that all of humanity both owned or could claim".

Whatever Venter's future now holds, it may be brighter than that of the company he created. Celera's share price has plummeted over the past two years from a March 2000 high of \$247 to \$15.

Celera last month abandoned its goal of selling genetic information, focusing instead on drug discovery (see *Nature* 416, 778; 2002). Analysts say that the free availability of genome databases in the public domain means that commercial ones are less lucrative than they were once touted to be. And, to judge by Celera's stock performance since its new business plan and president were announced on 22 April, investors are sceptical of its ability to compete in drug discovery. ■

Weapons lab's top job left unfilled

Rex Dalton, San Diego

The Lawrence Livermore National Laboratory in California is facing a leadership void after the University of California was forced to abandon its plans to appoint the facility's next director.

The university, which manages the nuclear-weapons laboratory for the Department of Energy, had planned to announce the appointment of Ray Juzaitis, an associate director of the Los Alamos National Laboratory in New Mexico, at a press conference on 26 April, sources say.

But university president Richard Atkinson cancelled the announcement after a last-minute intervention by the energy department. This dramatic reversal sparked fierce speculation in Washington and California about the delay. Atkinson had briefed Spencer Abraham, the energy secretary, a few days earlier, and received approval for

the candidate, a university spokesman says.

The fact that Juzaitis ran the nuclear-weapons physics division at Los Alamos — where Wen Ho Lee, the engineer accused and then acquitted of spying at the New Mexico laboratory, had worked — has been leaked to the press as a possible reason for the delay.

But the belated arrival of the Lee issue in the process is seen by some observers as a ruse by senior staff at Livermore to undermine the appointment of a director from Los Alamos, the other US nuclear-weapons laboratory, with which Livermore has always enjoyed a strained relationship.

University officials are now considering how best to proceed with the appointment. In addition to Juzaitis, sources say, three other candidates had been in contention for the job: Michael Anastasio and Jeffrey Wadsworth, both deputy directors at Livermore, and Steve Koonin, provost of the California Institute of Technology. ■

Air-travel study struggles to get airborne

Declan Butler, Paris

A partial rescue plan has been agreed for the first large-scale study of whether passengers on long-haul flights are at increased risk of developing life-threatening blood clots.

Scientists have obtained only around a quarter of the 12 million euros (US\$11 million) needed for the project, but they nonetheless aim to launch it this week, while continuing to search for the remaining funds.

The past month has seen a spate of lawsuits against airlines by victims of deep-vein thrombosis (DVT) — 'economy-class syndrome' — and their relatives. Proper scientific data are needed to improve our understanding of the risks involved so that we can offer sound safety advice for air travel, says William Toff, a cardiologist at the University of Leicester and a member of the project team.

The proposed project is far larger than previous studies of the problem. Its scientific protocol — called the WHO Research Initiative on Global Hazards of Travel (WRIGHT) — was approved by the World Health Organisation (WHO) last August.

The International Air Transport Association,



Up in the air: a proposed study would involve passengers on 200,000 flights — if funding can be found.

tion, which represents the airlines, has welcomed the study — at least in public — and has agreed to help the WHO and the International Civil Aviation Organization with the logistics involved.

Fund-raising is being done by the WHO,

but only 2.8 million euros has been raised so far, all from within the European Union. Frits Rosendaal, a researcher at Leiden University in the Netherlands, resigned from the project last month, citing a "lack of confidence in the complete commitment from the WHO".

Shanthi Mendis, a Geneva-based WHO official involved in project planning, says the WHO is "disappointed" that funding has not materialized. "We expected a more enthusiastic response from governments," she says. But several DVT researchers blame bureaucracy at the WHO for the lack of uptake.

At a meeting in Geneva last week of the project's scientific committee and its European funders, DVT researchers said that they would try to bypass the WHO and solicit funding directly from governments and other sources. To safeguard the study's independence, airlines will not be asked for money.

The meeting agreed that the first two phases of the study will begin this month. One will examine the role of passenger risk factors, such as alcohol consumption, seating conditions and predisposition to the disease. The second will involve the use of experimental pressurized cabins to determine whether low air pressure during flights contributes to the risks of sitting in cramped conditions for long periods.

In the final, most expensive phase, the study would track the health of passengers on some 200,000 flights. But this phase will begin only when full funding has been obtained. And the prospects for that have worsened following the decision by Australia — the nation that arguably depends most on long-haul flights — to snub the WHO effort and pursue its own study.

"We are all impatiently waiting for the funds, with our troops ready to go," says Bo Eklof, chief of the Straub Foundation's Vascular Center at the University of Hawaii.

Canadian lab loses Amgen backing

Jonathan Knight, San Francisco

A Californian biotechnology company is pulling the plug on a successful mouse-genetics institute at the University of Toronto.

The withdrawal of support from the Amgen Institute is the latest indication that biotechnology and pharmaceutical firms — faced with declining profits and empty drug pipelines — are losing some of their appetite for backing basic university research.

The six laboratories at the institute were considered to be exceptionally productive (see *Nature* 411, 519–520; 2001). But an agreement this week between Amgen, based in Thousand Oaks, California, and Toronto's University Health Network hands the administration of the institute to the Ontario Cancer Institute (OCI) in Toronto.

Although the financial details of the agreement have not been made public, Tak Mak, the Amgen Institute's director, says it will mean a sharp decrease in support from the Californian company. The institute's researchers will resort to individual grants for most of their funding, he says.

Since the institute's birth in 1993, its 50 or so scientists and technicians have been employed by Amgen, giving the company control of any intellectual property rights arising from their discoveries.

The move is part of a larger restructuring at Amgen, one of the oldest and largest

biotechnology firms, with 7,000 employees and sales last year of \$4 billion. The reshuffle is expected to align research more directly with product development.

Amgen will continue to provide support "at a level appropriate to the new arrangement", says Roger Perlmutter, Amgen's vice-president of research and development. It is estimated that Amgen's previous support amounted to around \$8 million per year.

Christopher Paige, vice-president of research at the OCI, says there are no plans to reduce the size of the research group. "We think the Amgen Institute is one of the best groups in the community," he says.



NASA faces legal challenge over Hawaiian telescope plan

Tony Reichhardt, Washington

A Hawaiian state agency is suing NASA over plans to upgrade the twin Keck telescopes on the mountain of Mauna Kea — the premier site for astronomical observation in the Northern Hemisphere.

The lawsuit, which was filed last week by the state's Office of Hawaiian Affairs, demands that the space agency completes a full Environmental Impact Statement (EIS) before carrying out its plan to build four to six small 'outrigger' telescopes that would improve Keck's ability to reveal planetary systems around distant stars. The case is seen as a significant escalation in a long-running battle between astronomers and native cultural groups in Hawaii.

NASA had hoped to be granted a state permit this month to begin construction later this year, and aimed to begin operating the outriggers in 2003. But completing an EIS would probably delay the project by a year or more, adding millions to its \$50-million price tag.

The outriggers would operate as an interferometer with the two existing 10-metre Keck telescopes. This effectively produces the resolution of a much larger instrument. NASA wants to upgrade the facility primarily to study planets outside our Solar System, and says that four of its six science objectives cannot be accomplished without the outriggers.

The agency had hoped that a less formal "environmental assessment" of the outriggers' impact on the mountain would satisfy native

Hawaiian groups, who have objected to the building of telescopes on a site that they hold sacred (see *Nature* **410**, 1015; 2001). In February, NASA proposed having cultural and archaeological monitors present during construction, and offered \$2 million in funding for cultural preservation and education projects in Hawaii. It also outlined a mitigation plan for the Wekiu bug, a threatened insect species that lives on Mauna Kea.

But the controversy "goes beyond the physical details to the spiritual aspects", says Bill Stormont, director of the Office of Mauna Kea Management, which was set up by the University of Hawaii to manage the science reserve on the mountain. And many astronomers fear that the NASA case is simply a symptom of a worsening relationship between themselves and native Hawaiian groups, which may ultimately threaten Mauna Kea's future as an observation site.

Although NASA's outriggers are the only significant construction planned on the mountain for the next several years, the University of California is eyeing the site for its California Extremely Large Telescope project. Mauna Kea is also considered a logical site for a possible Next Generation Large Telescope project.

But such plans are facing an increasing number of potential roadblocks. A bill introduced in the Hawaiian legislature in March, for example, called for astronomers to begin paying rent to use mountains. Although the proposal died during committee deliberations, another bill, which calls for a moratorium on telescope construction on Mauna Kea until environmental assessments have been completed, has already been passed by the state senate. ■

Neutrino review will seek out 'scientific redundancy'

Geoff Brumfiel, Washington

The White House has ordered a rapid-fire study of two proposed neutrino projects, raising the prospect of some tough choices ahead for the US physics community.

The projects to be reviewed are the IceCube neutrino observatory and the National Underground Science Laboratory (NUSL). Each would cost hundreds of millions of dollars to build, and both are devoted to the detection of neutrinos — tiny, almost massless particles that are of intense interest to physicists.

In a 29 March letter, John Marburger, director of the White House Office of Science and Technology Policy, asked the National Academy of Sciences for a study of the respective merits and of "any possible scientific redundancy" between the projects, to be completed by September.

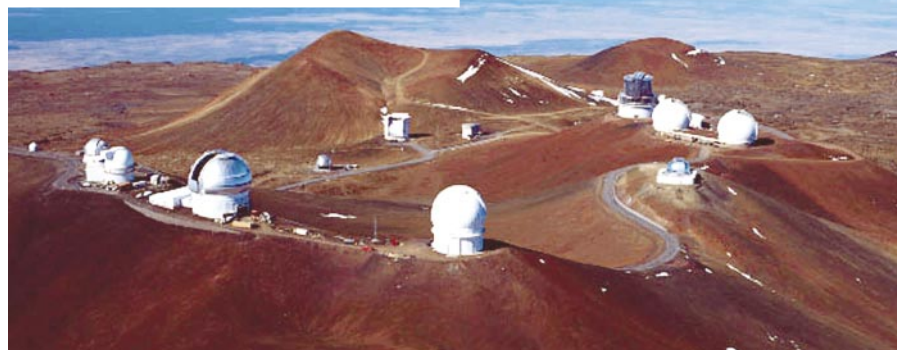
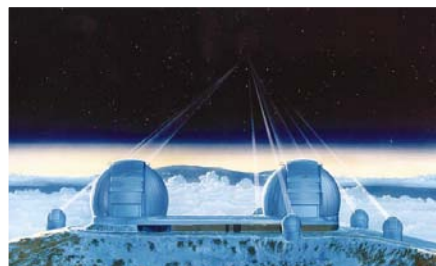
IceCube is an array of light-sensitive detectors buried within a cubic kilometre of Antarctic ice. The detectors are designed to watch for collisions between the ice molecules and high-energy neutrinos from deep space. From these collisions, astronomers hope to trace the neutrinos back to their extragalactic origins.

The international project, led by the National Science Foundation (NSF), has undergone several reviews — according to its leader, Francis Halzen of the University of Wisconsin at Madison, "they will have a hard time finding a person who has not reviewed it already".

The NUSL is new by comparison. Plans for its construction were submitted to the NSF only last June, and have yet to be endorsed by its governing National Science Board. But NUSL is riding a strong wave of scientific and political support (see *Nature* **416**, 775; 2002).

The underground laboratory would detect low-energy neutrinos and seek to determine their properties, as well as providing a home for proton-decay experiments, searches for 'dark matter' and geomicrobiology. "There is virtually no overlap in the physics of these two projects," says Wick Haxton of the University of Washington, who heads the NUSL plan.

But the NSF has a limited budget for major facilities, and budget hawks at the White House Office of Management and Budget, whose interest is said to have led to the study request, want to keep it that way. Project advocates hope the study will convince them that the two projects do not overlap. ■



NASA's proposed upgrade (inset) of Keck has provoked opposition from native Hawaiian groups.

NASA

UNIV. HAWAII INST. ASTRON., RICHARD WAINSCOTT/AP

Satellite delay spells shortage of star data

Alison Abbott, Munich

The recent decision to delay the launch of a German satellite has left astrometry — the study of the exact position of objects in space — facing a prolonged data drought.

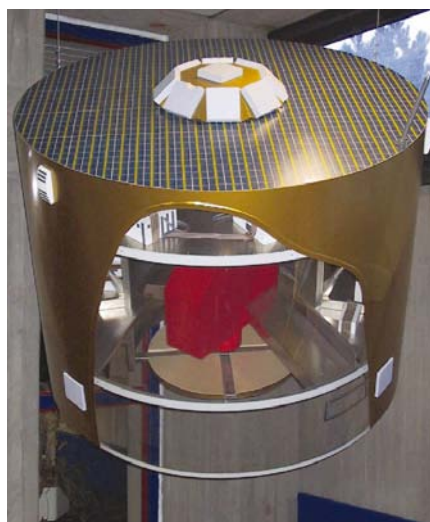
The satellite, known as DIVA, was scheduled to take off in 2004 on a mission to catalogue the positions of stars with unprecedented accuracy. But its launch has now been put back to at least 2007.

Astrometry specialists say that this delay — which closely follows the cancellation of NASA's FAME, a similar mission also scheduled to launch in 2004 (see *Nature* 415, 249; 2002) — will deprive them of useful data for nearly a decade. The next big astrometry mission, GAIA, which is being run by the European Space Agency (ESA), is expected to launch sometime between 2010 and 2014.

The astrometry community has been getting by on data collected between 1989 and 1993 by ESA's Hipparcos mission. DIVA and FAME were supposed to breathe new life into the community, attracting and training young researchers in preparation for the masses of data that GAIA will eventually deliver.

Both DIVA and FAME were designed to map tens of millions of stars, compared with the 100,000 contained in the Hipparcos catalogue. But GAIA should position as many as a billion stars, and is also expected to locate about 40,000 large planets outside the Solar System.

The catalogues prepared by such mis-



Grounded: a funding shortfall means that DIVA cannot be launched until 2007 at the earliest.

sions document the movement of stars in relation to each other, their brightness and their colour, as well as their position. Researchers can use this information to study cosmological questions — the relative motion of stars, for example, provides pointers to the age and mass of the Universe.

DIVA's main sponsor is the DLR, Germany's space agency, which had committed to paying half of the project's total cost of 57 million euros (US\$51.5 million). The rest of the money was to come from five German states, one of which pulled out, leaving a

funding gap of 15 million euros. "We saw no way to close the gap," says Sigmar Wittig, the DLR's director. "So we had no choice but to slow down the mission's development and plan a launch for 2007 or 2008."

GAIA, meanwhile, has substantially reduced its own projected costs by planning to use a Russian rather than a European launcher, which project scientists hope will help persuade ESA to fix a launch for 2010. ESA is to decide on a new scheduling programme for all its missions early this summer.

But an early launch for GAIA would render a late launch for DIVA redundant. The same redundancy might also scotch current efforts by US scientists to revive NASA's interest in FAME.

And failure to launch either of the smaller missions before the middle of this decade would damage the whole of astrometry, says Siegfried Röser of the Institute of Computational Astronomy in Heidelberg, and lead investigator for DIVA. Young scientists will steer clear of an area of research in which no new data are expected for years on end, he says.

This prospect has led some researchers to look at the possibility of combining DIVA and FAME into a single mission. "I'd look at collaboration very favourably," says Kenneth Johnston, of the US Naval Observatory in Washington DC, FAME's principal investigator. Röser says he is not averse to such a solution, but wants to see first if ESA can be persuaded to come to DIVA's aid. ■

No end in sight for German misconduct probe

Marion Kerstholt, Munich

An investigation into alleged scientific misconduct at a German university has fallen far behind schedule, prompting fears that the country lacks the mechanisms needed to handle such cases.

Officials at the University of Göttingen promised 10 months ago that the investigation would be completed by September 2001. But they now admit that they are no closer to releasing its findings.

The investigation concerns a clinical trial in which patients with kidney cancer were injected with a vaccine made up of tumour cells fused with cells from the immune system.

The researchers conducting the trial claimed that the treatment caused dramatic regression of secondary tumours (A. Kugler *et al. Nature Med.* 6, 332–336; 2000). But the trial was suspended last year after accusations of deception and careless clinical practice were levelled in the German media at two members of the team: Alexander

Kugler, then at the University of Göttingen, and Gernot Stuhler at the University of Tübingen (see *Nature* 412, 8; 2001). These claims are disputed by the researchers.

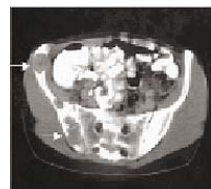
Horst Kern, president of the University of Göttingen, says that the investigation into the trial has been delayed primarily by data-protection rules covering patient information. Without access to the complete files of all 18 patients involved in the trial, the external task force looking into the allegations has been unable to proceed, he says.

"We find the situation unsatisfactory," Kern says, "but we do what we can to push things along." He declines to say when the investigation will be completed.

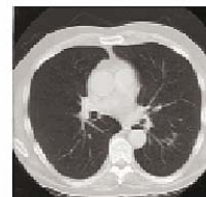
Clinical research in Germany has come under close scrutiny since 1997, when two biomedical researchers were accused of fabricating data in scores of papers.

Eberhard Hildt, a former postdoc at the University of Ulm, who first raised the alarm over the 1997 scandal, says it is in the interests of the scientific community for the

Before vaccination



After vaccination



Alleged irregularities halted a vaccine trial that had reported tumour regression.

latest investigation to be completed quickly. "Data-protection problems alone are no excuse," he says. "There seem to be no similar problems with patients' data when it comes to publishing a paper." ■

Seed imports raise hopes of Afghan recovery

Natasha McDowell, London

An international research consortium is taking a crucial first step in reviving the war-ravaged farms of Afghanistan by replenishing the country's seed stocks.

Led by the International Center for Agricultural Research in the Dry Areas (ICARDA), based in Aleppo, Syria, and involving the Afghan government as well as relief organizations, the consortium has already delivered 3,500 tonnes of wheat seed to almost 70,000 farmers in 11 provinces of Afghanistan in time for spring planting.

Three years of drought and more than 20 years of war have severely disrupted the country's agricultural system, and left seed stocks badly depleted. With two-thirds of Afghanistan's 26 million people living off the land, the success of the harvest is seen as fundamental to the stability and recovery of the nation. Restocking seed — especially wheat, which makes up 80% of the nation's grain production — is Afghanistan's first step on the path back to agricultural self-sufficiency.

"We need to ensure that these critical seeds survive," says Mohamed Sharif, deputy minister of agriculture of the Afghan Interim Administration. "Placing science at the forefront of our efforts will make this possible."

The revival effort is being funded by \$12 million in aid from the United States and Canada, and is orchestrated by Future Harvest, a network of 16 agricultural research centres, including ICARDA, that is supported worldwide by the Consultative Group on International Agricultural Research.

The consortium has used satellite imagery and soil maps to identify upland areas of the country that have sufficient soil moisture to stand a chance of yielding a rich harvest. It is installing new meteorological stations, called Agromet stations, to provide updates on ground conditions during the growing season. Each station measures moisture, soil temperature at three depths, and wind speed, and transmits the information via satellite to project administrators at Afghanistan's agriculture ministry and researchers at a new ICARDA office in Kabul.

Afghanistan is a diverse country in which knowledge of local agricultural conditions will be critical to the project's success, its participants say. But even under the Taliban regime, an extensive network of farmers kept working with local and international non-governmental organizations on seed trials. So the seed provided by the consortium includes drought-tolerant varieties that are known to be suited to local conditions.

The consortium has rushed the first supplies of improved seed into the country in time for the spring planting, so it can be grown and multiplied in time for a second planting in the autumn, when most of

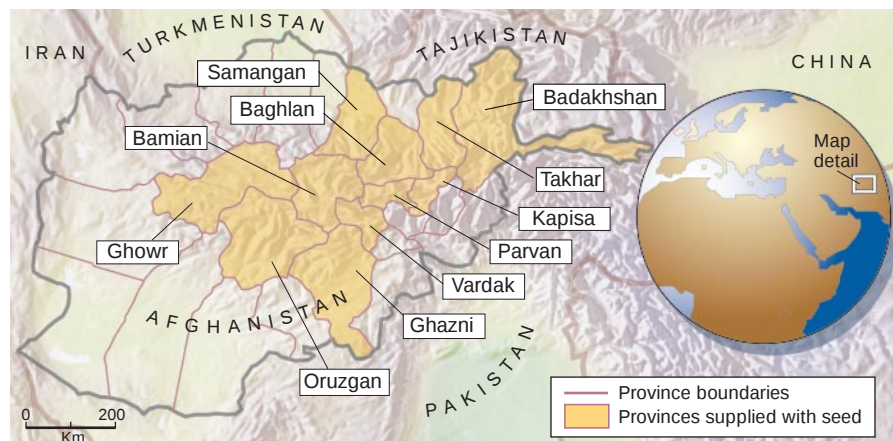
Afghanistan's harvest is grown, says John Dodds, a Washington representative of ICARDA. A further 60 tonnes of seed has been distributed to refugees passing through Kabul on their way home to remote parts of the country.

Farmers who buy the seed will receive guaranteed payments from the consortium for the spring harvest. The seed produced will be tested locally for quality and then redistributed before being sown in the autumn, along with a further 10,000 tonnes of imported seed — including some of the 2,000-plus indigenous Afghan seed varieties of wheat, barley and pulses that are already held by ICARDA.

"Satellite imagery and Agromet technology have been used before for relief efforts, but never on this scale," says Adel El-Beltagy, director general of ICARDA. The data generated will indicate which regions of the country will need food aid and, conversely, where aid supplies should be cut back. The challenge, Dodds says, is to phase out food aid so that it doesn't take the market away from farmers by depressing the price of locally produced grain.

The consortium is also comparing old satellite pictures of Afghanistan with new ones to assess damage to the country's once-extensive irrigation system. Traditional irrigation channels can be rebuilt at moderate cost, project officials say, but Soviet-built irrigation channels of half-open concrete piping will have to be abandoned and replaced.

Afghanistan was traditionally self-sufficient in food and had a strong agricultural research base, including an extensive crop gene-bank that operated until 1992. The consortium aims to supply 125,000 tonnes of seed over the next three years, and hopes that the nation will be self-sufficient again by 2007. This is a realistic goal, according to Dodds: "The consortium has a vision, but its feet are firmly on the ground."



Growth factor: seed imports to 11 provinces will help Afghanistan's farmers to become self-sufficient.

JASONS advice group set to work with defence directorate

Washington An independent group of scientists that advises the government on defence is poised to reach agreement with a new Pentagon sponsor.

The group, known as the JASONS, is a team of about 50 university researchers who have worked on some of the defence department's toughest technical challenges, including the ballistic-missile defence and submarine warfare. They are known for their secrecy and independence. The JASONS were dropped by their former sponsor, the Defense Advanced Research Projects Agency (DARPA), in March after the group refused to admit three agency-backed members into its ranks (see *Nature* **416**, 353; 2002).

The JASONS are now set to work directly with the Pentagon's directorate for Defense Research and Engineering, which oversees many of its research and development programmes — including DARPA. The contract should allow the JASONS to carry out this year's annual workshop without a hitch. They are likely to use the meeting to assess terrorist threats, such as the use of nuclear devices.

Germany outlaws stem-cell research while abroad

Munich German researchers have learned that they will be breaking the law if they work on new human embryonic stem-cell lines in foreign countries.

A new German law limits the cell lines that may be imported into the country to those created before 1 January this year (see *Nature*

415, 566; 2002). But other legislation forbids German scientists from using the cells in other countries. Germany's state secretary for research, Wolf-Michael Catenhusen, had wanted to amend the new law to make researchers immune to prosecution while working abroad, but was forced to back down before a second vote on the bill on 25 April.

"The worst thing is that we will not be able to collaborate with our European colleagues," says Oliver Brüstle, head of the Institute for Reconstructive Neurobiology in Bonn, who has championed the cause of human embryonic stem-cell research in Germany. "It is extremely serious and very unfortunate."

Dyke under scrutiny as protests flood in

Tokyo Researchers are being allowed to determine whether the controversial dyke at Isahaya Bay in southwestern Japan has damaged the local marine ecosystem — but they have been given just two months to do their work.

The dyke was built in 1997 as part of a plan to reclaim 700 hectares of land around the bay for farming. But seaweed farmers blame the dyke for falling seaweed harvests in the Ariake Sea into which the bay opens, and ecologists say that it may be responsible for a drop in the number of bottom-dwelling organisms in the bay itself.

The dyke was opened last week to give scientists two months to assess how it influences water quality and tides in the bay. But last year, an independent panel, endorsed by the previous cabinet's agriculture minister, recommended opening the dyke for a two-year study (see *Nature* **410**, 619; 2001). Ecologists have already complained about the new plans, saying



Current problem: researchers can investigate whether the Isahaya Bay dyke damages the ecosystem.

KYODO PHOTO SERVICE

that they have insufficient time to conduct a thorough study.

Pure science courses fail to entice French students

Paris New figures reveal a dramatic drop in the number of French students studying pure sciences at university, adding weight to fears that the country faces a research recruitment crisis.

The number of students studying physics at first-degree level has dropped by 46% since 1995, and the number of life-science students has fallen by 27%. The decline poses a threat to French science, as around half of all the country's researchers are due to retire over the next 10 years. Education minister Jack Lang has been instigating reform measures over the past few years, requiring schools to put more emphasis on project work and a hands-on approach to science, and asking universities to improve teaching methods and student supervision.

Other areas of French science education are in better health. The number of 16-year-olds specializing in science for their baccalaureate exam has remained more or less stable since 1995. And engineering and information-technology courses have shown a steady increase in entrants over the same period, as have more professionally orientated courses in science and technology.

British 'biobank' seeks public's genetic data

London The genetic details of half a million British people aged 45 to 69 are to be collected and offered to researchers under a new scheme aimed at pinpointing the causes of illness and disease. The genetic data will be stored alongside medical records and lifestyle information in a new UK 'biobank' and will be made available to both academic and commercial groups.

The £45-million (\$65-million) project was officially launched on 29 April by the bodies that are funding it — the Wellcome Trust, the Department of Health and the Medical Research Council (MRC). "This exciting project may one day herald a new era of medicine," says George Radda, chief executive of the MRC. "In 20 years' time we may see individualized approaches to disease prevention and treatment."

Nobel laureate solves taxing problem over earnings

Tokyo For a lucky and talented few, scientific awards can provide a useful addition to an academic salary. But as last year's joint winner of the Nobel Prize in Chemistry has just discovered, the taxman will want his share too.

Nagoya University chemist Ryoji Noyori



Quick reaction: Noyori has now paid the Japanese tax office the money he owed.

has been ordered by the Japanese tax office to pay some ¥15 million (US\$117,000) in back taxes and penalties on money earned during the seven years up to 1999. Noyori's untaxed income, which totalled ¥33 million, was earned from paid lectures at organic-chemistry societies around the world, and lucrative honours such as Saudi Arabia's King Faisal International Prize, which is worth US\$200,000. Noyori shared the 1999 King Faisal science award with Dieter Seebach, an organic chemist at the Swiss Federal Institute of Technology in Zurich.

Noyori, who received his Nobel Prize for studies of chiral molecules, has already paid the sum. He says that he was naive about the declaration of foreign-earned income, and that he did not know how to deal with the money as it was often bundled together with expense reimbursements.

Foreign firms fund French electronics

Paris French electronics research has received news of a billion-dollar investment from three major industrial groups.

French-Italian group STMicroelectronics, the Dutch company Philips and the US firm Motorola announced last month that they are to plough US\$1.4 billion over three years into a new microchip-fabrication facility and nanoelectronics laboratory in Crolles, near Grenoble. The money is thought to be the biggest single investment in French industry in the past 10 years.

The three companies, along with a fourth minor partner, the Taiwan Semiconductor Manufacturing Corporation, will focus the research component of the project on increasing the number of chips that can be made from a single silicon wafer, and the use of fine-etching technology to fit more transistors on a single chip. Academic research support will be provided by two Grenoble-based research centres — the Laboratory of Electronics and Information Technology, which is part of the French Atomic Energy Commission, and the National Polytechnic Institute, an engineering school.

Dual identities

Some people's blood contains cells from a sibling. Others are two individuals rolled into one. Yet more carry a distinct mutation in only parts of their bodies. Helen Pearson investigates chimaerism and mosaicism.

Eight years ago in Britain, a boy was born who, genetically, was two people. He was formed when two eggs, fertilized by two different sperm, fused into one embryo inside his mother's womb.

He was an unremarkable baby. But as a toddler, doctors discovered that he was a hermaphrodite — what was originally diagnosed as an undescended testis turned out to be an ovary, a fallopian tube and part of a uterus. Further investigation revealed that some parts of his body were genetically female but the rest, which contained a different combination of his parents' genes, was male¹.

The boy, who was otherwise healthy, is one of only a handful of known true human chimaeras — people carrying tissues that originated in two separate embryos. More common are mosaics, who have patches of tissue that differ genetically from the rest of their body, thanks to a mutation or chromosomal anomaly that arose early in embryological development.

The frequencies of chimaerism and mosaicism are unknown, but doctors might benefit from a better understanding of both conditions. In recent years, tantalizing hints have emerged that pockets of genetically mismatched cells may contribute to conditions as common as infertility, autism and Alzheimer's disease. "I think mosaicism has been neglected as an underlying cause of disease," says Huntington Potter, who works on the genetics of Alzheimer's at the University of South Florida in Tampa.

And if chimaeras and mosaics are more common than we realize, they will complicate future efforts to tailor drug treatments to people's individual genetic constitutions. Two genetically different tissues in one body might produce an unpredictable

response to a drug, speculates Roland Wolf, who studies pharmacogenetics at the University of Dundee, UK. "It's completely unknown," he says.

The twin within

Human chimaerism first came to light with the advent of blood typing — some people, it emerged, have more than one blood group. Most are 'blood chimaeras', non-identical twins who shared a blood supply in the womb. Those who were not born a twin are thought to be pumping around the remnants of a sibling that died early in gestation and was spontaneously aborted. One British woman, for instance, was

unaware that she once had a twin until routine blood tests during her pregnancy in the early 1980s revealed a population of chromosomally male blood cells².

Twin embryos often share a blood supply in the placenta, allowing blood stem cells to pass from one embryo and settle in the bone marrow of the other, seeding a lasting source of blood. As a result, as many as 8% of non-identical twin pairs have chimaeric blood³. And given that most multiple conceptions that result in live births involve the loss of one twin early in pregnancy⁴, there may also be significant numbers of blood chimaeras among single births.

Even more people have 'micro-chimaerism', carrying smaller numbers of foreign blood cells that may, for instance, have passed between mother and fetus across the placenta, or persist from a blood transfusion. Some researchers argue that the presence of foreign white blood cells might help to explain autoimmune diseases, in which the immune system turns on the body's own tissues⁵.

True chimaeras, in which many tissues are affected, are thought to be very rare, and can form when non-identical twin embryos fuse shortly after fertilization. "If you've got two embryos there's the chance of two becoming one," says clinical geneticist David Bonthron, who led the team at the University of Edinburgh, UK, that reported on the British hermaphrodite boy.

Chimaerism affecting a variety of tissues can also result from other events. In 1995, for instance, Bonthron described another boy who was partially parthenogenetic: cells from his blood and certain other tissues contained none of his father's chromosomes; instead, they featured a duplicated set of one half of his mother's⁶. Although it is not



Cryptic conditions: chimaerism and mosaicism are often only spotted by chromosomal analysis.



Lines of inquiry: mosaicism causes these unusual pigmented patterns called Blaschko's lines.

J. KING-HOLMES/SPL

R. HEPPLE

unknown for an egg to start developing without being fertilized, fully parthenogenetic human embryos cannot develop to term. Bonthron, now at the University of Leeds, UK, believes that the partially parthenogenetic boy owed his unusual genetic constitution to an egg that spontaneously divided into two cells, one of which was fertilized. The second cell then copied its maternal chromosomes, allowing the resulting chimaera to form a viable embryo.

Hard to spot

True chimaeras, including both of those identified by Bonthron's team, generally only come to light if they contain male and female cells, when they can cause hermaphroditism or a mismatch between a person's sexual organs and their chromosomal sex as revealed by a blood test. So might the condition be more common than we realize? "I'm convinced that on the streets of London and Hamburg there are many undetected chimaeras," says Rudolf Happle, a dermatologist at the University of Marburg, Germany, who has long been fascinated by mosaicism and chimaerism.

The rise of *in vitro* fertilization (IVF) is almost certainly bringing more chimaeras into the world. To improve success rates, two or more embryos are placed in the uterus, which explains why women undergoing IVF have up to 25% more twin pregnancies than usual. More twins means more chimaeras, says Bonthron, who notes that the

British hermaphrodite boy was an IVF baby¹.

Mosaicism is more common than chimaerism and is also better studied. Human mosaics arise when a mistake during cell division in the early embryo stops the correct number of chromosomes segregating to each cell, or creates a mutation in a single gene. If this happens in one of the first few cell divisions after fertilization, a large proportion of cells will inherit the defect.

Patchy diseases, in which only regions of tissue are affected, might be caused by mosaicism. Another telltale sign of the condition is a characteristic variation in skin pigmentation that causes patterns called Blaschko's lines, including V-shaped streaks on the back that are sometimes only visible under ultraviolet light.

But pinning down the specific mutation in a mosaic disease can prove tricky. Leslie Biesecker of the National Human Genome Research Institute in Bethesda, Maryland, is attempting to identify the gene that is mutated in Proteus syndrome. This is the disfiguring condition that the Victorian era 'Elephant Man', Joseph Merrick, is thought to have suffered from. The syndrome's symptoms of patchy tissue overgrowth have led researchers to speculate that it is caused by a mosaic mutation.

Threads and patches

Biesecker hopes to compare the genes that are active in diseased patches of tissue with those in unaffected areas using DNA microarrays. But obtaining tissue samples from the 100 or so sufferers of Proteus syndrome worldwide is difficult, and the genetic differences between tissues may be slight.

Other researchers suspect that mosaicism could be involved in more common diseases. Wendy Robinson of the University of British Columbia in Vancouver, for instance, is intrigued by the observation that the placenta becomes mosaic in about 2% of pregnancies. Often, these mosaic placentas contain patches of cells that have an extra chromosome — a condition called trisomy. Because both fetus and placenta develop from the same cells, Robinson wonders if many fetuses contain undetected patches of trisomic tissues that persist into adult life. "People could have little pockets of trisomic cells sitting around inside



David Bonthron suspects that fertility treatments are increasing the number of human chimaeras.

them that later lead to disease," she suggests. Intriguingly, her team has found that some women who experience recurrent miscarriage carry trisomic cells⁷.

Other researchers theorize that a hidden patch of brain cells with an extra copy of chromosome 21 could be what predisposes some people to Alzheimer's disease. This idea arose from the long-standing observation that people with Down's syndrome, who carry an extra chromosome 21 in all of their cells, develop symptoms of Alzheimer's at an early age. Recently, two research groups, including the South Florida team led by Potter, have found that many patients with Alzheimer's also have low levels of cells with an extra copy of chromosome 21 cells circulating in their blood^{8,9}.

Autism is also being explored for links with mosaicism. In unpublished research, Susan Folstein of Tufts University School of Medicine in Boston examined autistic children under ultraviolet light and found that as many as 10% of them had pronounced Blaschko's lines. She suspects that a mosaic patch of brain cells, unable to migrate or communicate with its neighbours, might be what causes some cases of autism. But until an underlying mutation can be found, the idea remains unproven.

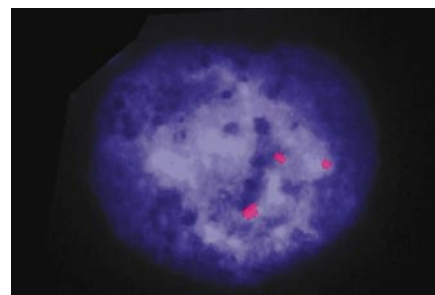
Given the intriguing results that are starting to emerge, researchers who have studied chimaerism and mosaicism are keen to spread the word so that more attention is paid to the clinical significance of the conditions. At present, they say, most doctors and clinical geneticists are simply not looking. ■

Helen Pearson works in Nature's news syndication team.

1. Strain, L., Dean, J. C. S., Hamilton, M. P. R. & Bonthron, D. T. *N. Engl. J. Med.* **338**, 166–169 (1998).
2. Bird, G. W. *J. Immunogenet.* **9**, 317–322 (1982).
3. van Dijk, B. A., Boomsma, D. I. & de Man, A. J. M. *Am. J. Med. Genet.* **61**, 264–268 (1996).
4. Boklage, C. E. *Int. J. Fertil.* **35**, 75–94 (1990).
5. Nelson, J. L. *Trends Mol. Med.* **8**, 109–113 (2002).
6. Strain, L., Warner, J. P., Johnston, T. & Bonthron, D. T. *Nature Genet.* **11**, 164–169 (1995).
7. Sangha, K. K., Stephenson, M. D., Brown, C. J. & Robinson, W. P. *Am. J. Hum. Genet.* **65**, 913–917 (1999).
8. Migliore, L. *et al. Cytogenet. Cell Genet.* **87**, 41–46 (1999).
9. Geller, L. N. & Potter, H. *Neurobiol. Dis.* **6**, 167–179 (1999).

I think mosaicism has been neglected as an underlying cause of disease.

Huntington Potter



Some Alzheimer's patients have three copies of chromosome 21 (red dots) in certain cells.

Snowball fights

Did the world freeze over some half a billion years ago? Two Harvard scientists think so, but convincing other climatologists is proving difficult. Naomi Lubick tracks the latest twists and turns in the snowball Earth debate.

Paul Hoffman and Daniel Schrag have had a busy few years. In 1998, the two Harvard University geologists rekindled a radical idea: that on at least one occasion between 580 million and 750 million years ago, the Earth lay entirely encrusted in ice for tens of millions of years. This 'snowball Earth' hypothesis seemed to explain some puzzling geological data. But it was controversial then, and the debate shows no sign of letting up.

Sceptics first asked how the Earth could freeze and thaw in such a short geological time. Climate modellers have since questioned whether ice sheets could have reached the Equator. And last year came an assault on Hoffman and Schrag's central line of geological evidence. The proponents of snowball Earth, it seems, are on the defensive once more.

The idea of a global glaciation was first proposed in the 1960s by Mikhail Budyko of the Main Geophysical Observatory in St Petersburg, Russia. Budyko looked at what would happen if the Earth's climate were to cool slightly, prompting an increase in the size of the polar ice-caps. Ice reflects heat from the Sun, so this growth would cause further cooling. Runaway growth of the ice-caps could result, Budyko argued, eventually leaving the Earth entirely sheathed in ice¹.

Budyko's ideas explained puzzling evidence, including signs of scouring of rocks by ice, that seemed to imply that glaciers reached the Equator on at least two occasions between 580 million and 750 million years ago, towards the end of the Neoproterozoic period. This was baffling, because ice sheets reached only as far as northern Europe dur-

ing more recent ice ages. But Budyko's theory had some holes in it. What, for example, eventually caused the ice to thaw?

Iron out

In 1992, Joseph Kirschvink, a geologist at the California Institute of Technology in Pasadena, provided an explanation of how the ice could have receded². Kirschvink, who coined the term 'snowball Earth', realized that normal cycles of rain and erosion, which play an important role in removing carbon dioxide from the atmosphere, would have shut down if ice had covered the oceans. Carbon dioxide released by volcanoes would then build up in the atmos-



Volcanic CO₂ may have caused a greenhouse effect that freed snowball Earth from its ice age.

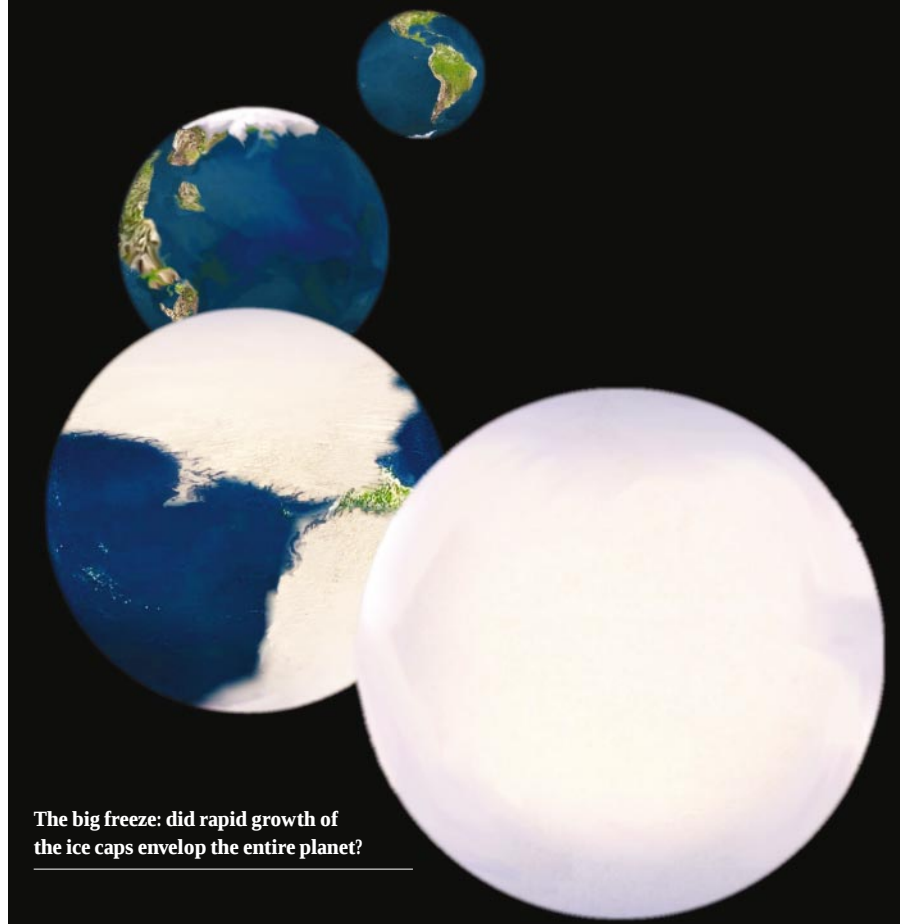
phere, eventually creating enough greenhouse warming to melt the ice sheets.

Kirschvink also pointed out that a snowball Earth could explain another strange geological deposit — iron-rich rocks that formed near the end of the Neoproterozoic. Iron is added to the ocean at geothermal vents in the sea floor and precipitates out of sea water when it comes into contact with oxygen. But if the oceans had been capped with ice, oxygen levels in water would have fallen and dissolved iron would have built up. Oxygen levels would have increased when the ice melted, causing large amounts of iron to precipitate out and fall to the sea floor.

Six years later, Hoffman and Schrag, together with colleagues at Harvard, published the paper that thrust the hypothesis back into the limelight³. They had studied ratios of carbon isotopes in rocks formed when carbon-containing compounds precipitated out of sea water. Photosynthetic marine microorganisms take up carbon, preferring the lighter carbon-12 isotope to the heavier carbon-13 — so photosynthesis causes carbon-12 levels in water to fall, leaving less of that isotope to precipitate out.

But when Hoffman and Schrag looked at 'cap carbonates' — sediments that were deposited towards the end of the Neoproterozoic glaciations — they found surprisingly high levels of carbon-12. In fact, the ratio of carbon isotopes suggested that almost no photosynthesis had occurred in the waters from which the rocks precipitated. This, they reasoned, was exactly what would occur if ice had covered the ocean and starved it of light.

Journals' correspondence columns were



The big freeze: did rapid growth of the ice caps envelop the entire planet?

soon buzzing with debate over the paper. Some critics pointed out that there was no evidence for a lowering of sea level that would be expected to accompany the freezing⁴. Hoffman countered by disputing the reasoning behind claims that sea level should fall⁵. The pattern was set: critics have continued to assail the snowball Earth theory, while Hoffman and, to a lesser extent, Schrag have defended the idea. With some of the most serious criticisms of their idea arising in the past year, the pair are as busy as ever.

Scouring the evidence

Most researchers now accept that ice reached equatorial regions — but only on land. Evidence for this comes from the deposits that once puzzled geologists. The alignment of magnetic particles in some rocks can be used to determine the latitude at which they formed, as the Earth's magnetic field — which pulls the particles into line as the rock forms — is inclined at different angles at different latitudes. Some Neoproterozoic rocks with equatorial magnetic signatures appear to show signs of scouring from moving ice.

But Michael Arthur, a geochemist at Pennsylvania State University in University Park, who is sympathetic to the snowball Earth theory, says that there still is not enough evidence to demonstrate that the tropical oceans froze over. "The nature of the glaciation is not well known," he says. "Is it a big ice sheet? Or are they just mountain glaciers coming down?"

Others share his doubts. In January this year, geologist Daniel Condon and colleagues at the University of St Andrews, UK, published an analysis of rocks formed on the sea floor during the period when Hoffman and Schrag claim the oceans were iced over. These rocks contained layers of stony debris of the kind carried by glaciers⁶. If the debris reached the sea floor, then the ice must have melted. And the layers are so close together, claims Condon, that the oceans must have been

unfrozen for at least some time each year.

Climate modellers are also reluctant to embrace snowball Earth. "It's very, very difficult to simulate," says Chris Poulsen, a modeller at the University of Southern California in Los Angeles. Last year, he published a simulation showing that the ice sheets would have stopped at northern Europe during the late Neoproterozoic⁷. One problem, say the modellers, is that oceans contain too much heat for them to freeze over completely.

Even the evidence from cap carbonates, which is considered by Hoffman and Schrag to be the 'smoking gun' for snowball Earth, has recently come under attack. Last May, Martin Kennedy, a palaeoclimatologist at the University of California, Riverside, proposed that the unusual carbon isotope ratios in the cap carbonates were caused by a sudden release of methane gas⁸.

Methane matter

Kennedy, together with Nicholas Christie-Blick and Linda Sohl of the Lamont-Doherty Earth Observatory in Palisades, New York, noted that large amounts of organic matter would have been trapped under ice sheets on land during the Neoproterozoic. Methane released when this matter decomposed would have remained trapped in this ice. When the ice started to melt, ice sheets close to the oceans would have been flooded by rising sea levels, and the methane slowly released. But this methane, like the organic matter that produced it, would have contained high levels of carbon-12 and could have caused the skewed isotope ratios seen in the cap carbonates.

Last December, Kennedy and Christie-Blick teamed up with Anthony Prave of the University of St Andrews to publish a paper that cast doubt on another piece of evidence central to the snowball Earth hypothesis. In an analysis of samples of carbonate rocks from Namibia that formed before the melt-



Chris Poulsen's simulation of temperatures on Neoproterozoic Earth suggest that polar ice sheets (white lines) did not reach low latitudes.

ing of the land ice⁹, they found little evidence of the odd isotope ratios that Hoffman and Schrag found in their cap carbonates. "The results indicate a very normal ocean," says Kennedy. "There was clearly no ice sheet."

Hoffman and Schrag have not taken such criticisms lying down, and their rebuttals are scattered throughout the geological literature of the past few years. In March this year, for example, Hoffman took on what he described as Kennedy's "creative" methane hypothesis¹⁰.

Kennedy had pointed to several unusual features of cap carbonates, such as tube-shaped cavities, as evidence that methane gas was released into the rocks as they were formed. But Hoffmann counters that these features are absent from other cap carbonates in the same area, and also disputes whether there would have been sufficient organic matter to generate the methane. To the modellers, he points out that the oceans have reached freezing point during the coldest periods of the past 10,000 years. Hoffman, sometimes with Schrag, has managed to publish a reply to almost every article on snowball Earth.

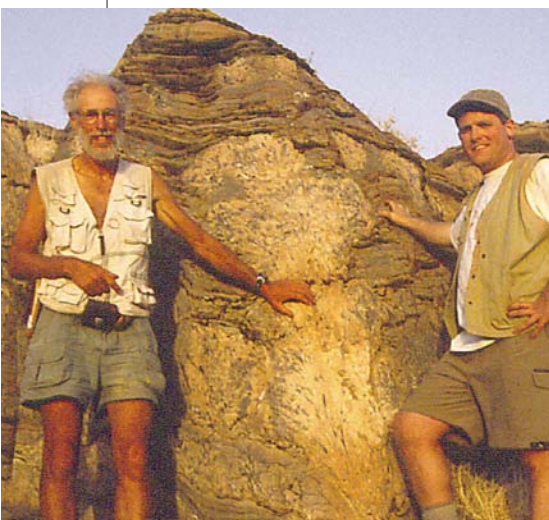
Although the debate remains polarized, most geologists and climate researchers believe that elements of many of the competing ideas are likely to be correct. But few palaeoclimatologists back the 'hard' version of snowball Earth, with its ice-covered oceans. A 'slushball Earth', with oceanic ice sheets extending to middle latitudes, seems to be gaining broader support.

But Hoffman remains unperturbed, arguing that radical ideas always face a barrage of scepticism. Both plate tectonics and evolution by natural selection faced similar battles for acceptance, for example. "I like to say I'm Kirschvink's bulldog," he laughs. ■

Naomi Lubick is a freelance writer in Palo Alto, California.

1. Budyko, M. I. *Tellus* **21**, 611–619 (1969).
2. Kirschvink, J. L. in *The Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 51–52 (Cambridge Univ. Press, New York, 1992).
3. Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. *Science* **281**, 1342–1346 (1998).
4. Williams, G. E. *Nature* **398**, 555–556 (1999).
5. Hoffman, P. F. *Nature* **400**, 708 (1999).
6. Condon, D. J., Prave, A. R. & Benn, D. I. *Geology* **30**, 35–38 (2002).
7. Poulsen, C. J., Pierrehumbert, R. T. & Jacob, R. L. *Geophys. Res. Lett.* **28**, 1575–1578 (2001).
8. Kennedy, M. J., Christie-Blick, N. & Sohl, L. E. *Geology* **29**, 443–446 (2001).
9. Kennedy, M. J., Christie-Blick, N. & Prave, A. R. *Geology* **29**, 1135–1138 (2001).
10. Hoffman, P. F., Halverson, G. P. & Grotzinger, J. P. *Geology* **30**, 286–287 (2002).

D. SCHRAG



Paul Hoffman (left) and Daniel Schrag claim that snowball Earth left its hallmark on rocks...



...but Martin Kennedy argues that the sediments' carbon signature was left by a methane release.

M. J. KENNEDY

Information overload

The tide of genetic data threatening to swamp researchers has led a 'data warehousing' firm to tune in to science. Carina Dennis charts its move from airlines and banks to biology.

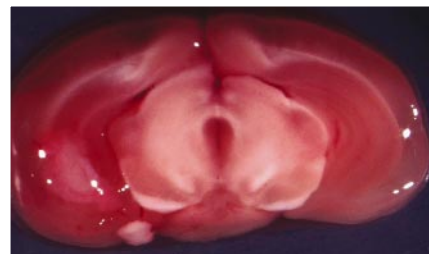
What do biologists interested in gene expression have in common with airlines, banks and food retailers? Data, data ... and yet more data. A range of high-throughput techniques such as DNA microarray analyses are providing biologists with sets of data that dwarf anything they have ever dealt with before. And in this rising tide of information, a company that supplies data 'warehouses' has spied an opportunity.

Although largely unknown in scientific circles, Teradata of San Diego is a market leader in supplying systems that can manage more than a terabyte (1 trillion bytes) of data. Its databases help to organize supermarkets' supply chains so that shelves get filled, for instance, and are used to keep track of airline reservations and banks' financial records.

Teradata's excursion into biology follows an introduction to Carolee Barlow, a neurobiologist at the Salk Institute for Biological Studies in La Jolla, California. Barlow and David Lockhart, president of Ambit Biosciences in San Diego, have a grant from the US National Institutes of Health to generate gene-expression profiles of the mouse brain. Monitoring the activity of thousands of genes in more than 150 different cell types in several



Checking in: a powerful database used to track airline reservations will soon be tracing gene expression in mouse brains (right).



mouse strains and under a range of different experimental conditions means that the project will generate about three-quarters of a terabyte of data. "It became clear that most commercially available databases were not suited to what we needed," says Lockhart.

The idea of linking up with Teradata came through a chance conversation between Barlow and Sudhakar Shenoy, chairman and chief executive of Information Management Consultants (IMC) of McLean, Virginia, a company that develops specialized software tools and databases for a number of clients including the US government. "Banks, airlines and Walmart have much bigger problems with data than us," says Barlow. "When we told Teradata of our problem, it was a no-brainer for them."

Pump up the volume

The result was a collaboration between the biologists, IMC and Teradata to adapt the latter's data warehouse for large-scale gene-expression analyses. Lockhart and Barlow supply the data and devise algorithms to ask relevant questions of the database; IMC and Teradata provide database infrastructure and software engineering. The system is expected to be ready within a month and Barlow says that trials have allowed analyses that previously took days to be done within hours.

Once the system has been fully tested, Teradata plans to target the life-sciences market. "We think the bioresearch sector is a good potential market for Teradata technology," says Alan Chow, vice-president and general manager of Teradata's development division. At a cost of about US\$1 million for its most basic terabyte data warehouse, buying a Teradata system is beyond the pocket of most biology labs — although pharmaceutical companies might be interested. But IMC hopes to develop an outsourcing service,

enabling biologists with particularly difficult data-management problems to use Teradata systems for a reasonable subscription fee.

Bioinformaticists are uncertain about Teradata's prospects in the commercial marketplace for biological databases, where companies such as Oracle and IBM are established players. With relatively few of today's repositories for biological data operating in the terabyte realm, Lincoln Stein of the Cold Spring Harbor Laboratory on Long Island, New York, thinks the demand for high-end data warehouses may be limited — a view echoed by Shankar Subramaniam of the University of California, San Diego. "Teradata is an excellent data warehouse for large volumes of data — the larger the volume of data, the better it is," says Subramaniam. But he argues that its specialized hardware and software makes it less suitable for users who do not need such enormous capacity.

Jeff Augen, director of strategy with IBM Life Sciences in Somers, New York, adds that bioinformatics often involves relating information in disparate formats held in different places. "Very often you have to link many databases together and the Teradata system was not designed to be used in that way," he argues.

But Richard Winter of the Winter Corporation, a database consultancy firm in Waltham, Massachusetts, suspects that other biologists will find themselves in Barlow's position, deluged with more data than their current systems can manage and analyse. "Teradata offers a solution to a class of problems which may be important to the life-sciences sector over the next few years," he says. "Biologists are working their way up the learning curve of how they can exploit database technology."

Carina Dennis is *Nature's* Australasian correspondent.

♦ www.teradata.com



Carolee Barlow and David Lockhart expect to generate three-quarters of a trillion bytes of data.

Don't punish scientists for government actions

Sir — You mention in News in Brief (*Nature* **416**, 574; 2002) a letter from 125 academics published in the UK national newspaper *The Guardian*, urging a suspension in European funding for Israeli research programmes.

Richard Dawkins, a signatory of that letter, once notoriously advised the Prince of Wales: "Of course we must be open-minded, but not so open-minded that our brains drop out." With this in mind, I would like to point out to the 125 distinguished signatories that Israel is not a freeloading recipient of European research funds, but rather a participating country that pays a designated contribution to the overall budget of the European Union (EU).

A suspension of our 'associated nation' status in the EU will free our government from the obligation to transfer these funds to the EU. Because academic research and academic freedom are not priorities of the Sharon government, to put it mildly, the net result will be a punitive blow to Israeli researchers (not a group noted for blind support of the government), while freeing up funds that the Sharon government will undoubtedly channel into defence needs.

Perhaps the 125 *Guardian* authors need to bear in mind another (slightly modified) quotation, "put brain into gear before setting pen to paper".

Mike Fainzilber

Department of Biological Chemistry, Weizmann Institute of Science, 76100 Rehovot, Israel

The challenge offered by X-ray lasers

Sir — In his Correspondence "Excitement over X-ray lasers is excessive" (ref. 1), responding to your News Feature² about the applications of X-ray free-electron lasers (FELs), Richard Henderson raises two points. First, that the development of X-ray FELs for biological applications threatens the field of structural biology because it will use up money needed for existing techniques; and second, that the development of existing techniques will be more fruitful than investing in an expensive, unproven technology such as X-ray FELs. I believe that both of these arguments are incorrect.

Synchrotron machines were developed by physicists and materials scientists, and have evolved to benefit biology. When the first dedicated synchrotron light sources came to life more than 20 years ago, they produced X-rays with intensities

50–100 times greater than a rotating anode generator. Yet synchrotron radiation triggered a revolution in the life sciences, and today almost 15,000 X-ray structures feature in the protein data bank as a result.

Did the construction of synchrotrons harm biology? Certainly not. Should one be afraid of a much bigger step ahead when the first X-ray FEL comes to life? I don't think so.

If one takes the speed of walking and multiplies it by 10^{10} , the result is a speed 100 times faster than the speed of light. The expected improvement in peak brilliance offered by X-ray FELs over existing synchrotrons is of this magnitude. These machines will bring us into a world for which only predictions exist.

A lack of hands-on experimental data on the behaviour of matter under these conditions makes very detailed forecasts difficult. But one thing is certain: X-ray FELs will generate fresh thinking, new science and a new scientific community. To imply that there is nothing that X-ray FELs can discover in biology is misguided.

The money for building and running the Linac Coherent Light Source at Stanford (first of the planned X-ray FELs)

will come from the US Department of Energy. These funds have no direct 'crossover' to biomedical research or biology. Every one of the five synchrotron X-ray sources in the United States was built and is operated by agencies that do not deal with biology.

One can make arguments for the advantages of using electrons, as outlined by Henderson, or others for using X-ray pulses. One can estimate that the conventional handicap of X-rays over electrons could be reversed³ and made into a net gain on very small samples, when extremely intense and very short X-ray pulses will be produced. This is the domain of the X-ray FELs.

There is surely room for more than one avenue to be explored — different techniques will be appropriate for different biological structures and processes. This flexibility is intrinsic to the nature of research and essential for scientific progress. There are no threats here, only challenges.

Janos Hajdu

Biomedical Center, Uppsala University, Box 576, S-751 23 Uppsala, Sweden

1. Henderson, R. *Nature* **415**, 833 (2002).

2. Patel, N. *Nature* **415**, 110–111 (2002).

3. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. *Nature* **406**, 752–757 (2000).

Intensive farming, US-style, is not sustainable worldwide

More greenhouse gases will increase loss of usable land.

Sir — Stephen Badiansky in his Correspondence¹ "How affluence could be good for the environment" calculates that, thanks to increased crop yields, the 'ecological footprint' of North Americans is much smaller than that calculated by the World Wide Fund for Nature. He touches on the inclusion of per capita greenhouse-gas emissions, but he overlooks the impact of intensive crop production on emissions of the non- CO_2 greenhouse gases methane (CH_4) and nitrous oxide (N_2O).

Agricultural soils and livestock farming already emit around 7 million tonnes of nitrogen as N_2O to the atmosphere each year²; worldwide nitrogen-fertilizer applications at the rate Badiansky suggests as typical for the United States (115 kg ha^{-1}) would lead to further increases. Similarly, agriculture around the world constitutes a source of more than 100 million tonnes of methane each year — more intensive farming is likely to increase this methane source, while at the

same time increased global nitrogen application and deposition may greatly reduce the soil methane sink (about $30 \text{ Tg CH}_4 \text{ yr}^{-1}$; ref. 3).

Badiansky argues that, with more intensive farming methods and as countries become wealthier, the world's population could significantly increase per capita consumption without increased land requirements.

I doubt whether he has included in his calculations the loss of usable land owing to sea-level rise, as predicted by the Intergovernmental Panel on Climate Change, and to desertification resulting from global warming⁴.

David S. Reay

Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH6 3JU, UK

1. Badiansky, S. *Nature* **416**, 581 (2002).

2. Kroeze, C. A., Mosier, A. R. & Bouwman, L. *Global Biogeochem. Cycles* **13**, 1–8 (1999).

3. Mosier, A. R. et al. *Climatic Change* **40**, 39–80 (1998).

4. IPCC. *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).

The discipline will have to reinvent itself if it is to survive and flourish.

H. Charles J. Godfray

Taxonomy, the classification of living things, has its origins in ancient Greece and in its modern form dates back nearly 250 years, to when Linnaeus introduced the binomial classification still used today. Linnaeus, of course, hugely underestimated the number of plants and animals on Earth. As subsequent workers began to describe more and more species, often in ignorance of each others' work, the resulting confusion and chaos threatened to destroy the whole enterprise while still in its infancy. In today's jargon, we might call this the first bioinformatics crisis. Using the tools then available, nineteenth-century taxonomists solved this crisis in a brilliant way that has served the subject well since then. They invented a complex set of rules that determine how a species should be named and associated with a type specimen; how generic and higher taxonomic categories should be handled; and how conflicts over the application of names should be resolved. All these rules revolved around publications in books and scientific journals, and their descendants form the current codes of zoological and biological nomenclature.

But today much of taxonomy is perceived to be facing a new crisis — a lack of prestige and resources that is crippling the continuing cataloguing of biodiversity. In the United Kingdom, a Parliamentary Select Committee is currently conducting an enquiry into the health of the subject for the second time in 10 years, and similar concerns are being expressed around the world. In this article I shall first explore why descriptive taxonomy is in such straits (in contrast, its sister subject, phylogenetic taxonomy, is flourishing). Then, after this essentially negative exercise, I will argue that taxonomy can prosper again, but only if it reinvents itself as a twenty-first-century information science. It needs to adopt some of the solutions that molecular biologists have developed to cope with the second bioinformatics crisis: the huge explosion of sequence, genomic, proteomic and other molecular data.

The problem

Why can't descriptive taxonomy attract large-scale funds in the same way as other big programmes like the Human Genome Project or the Sloan Digital Sky Survey? All three projects are enabling science: not in themselves generating new ideas or testing hypotheses, but allowing many new areas of research to be opened up.

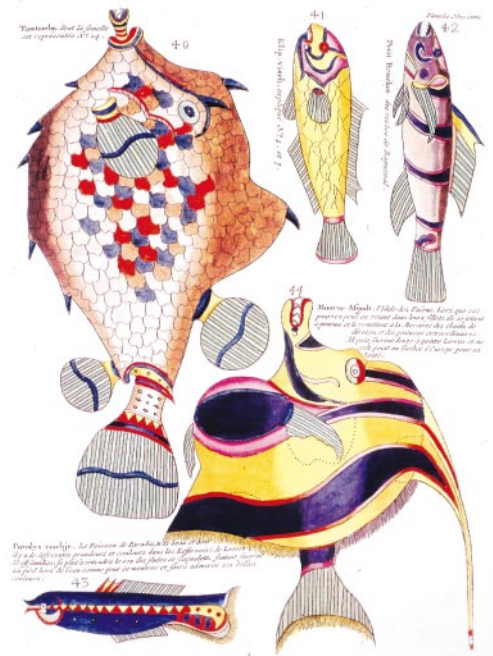
One reason is that taxonomists lack clearly achievable goals that are both realistic and relevant. Of course it would be great to describe every species of organism on Earth, but we are still monumentally uncertain as to how many species there are (probably somewhere between 4 million and 10 million); this goal is just not realistic at present. There are various projects aimed at listing, for example, all the valid described species of animal in Europe, or butterflies on Earth (see Box 1, overleaf). These aims are eminently achievable and very worthwhile, but the results are like raw, unannotated DNA sequences: unexciting and of relatively little value in themselves to non-specialists. Taxonomists need to agree on deliverable projects that will receive wide support across the biological and environmental sciences, and attract public interest.

A second problem is part of the legacy of more than 200 years of systematics. Many taxonomists spend most of their career trying to interpret the work of nineteenth-century systematicists: deconstructing their often inadequate published descriptions, or scouring the world's museums for type material that is often in very poor condition. A depressing fraction of published systematic research concerns these issues. In some taxonomic groups the past acts as a dead weight on the subject, the complex synonymy and scattered type material deterring anyone from attempting a modern revision. As Frank-Thorsten Krell pointed out in Correspondence (*Nature* 415, 957; 2002), "original descriptions have to be referred to for ever, independent of the paper's quality".

The problems do not always lie in the past. Even today, many species are being described poorly in isolated publications, with no attempt to relate a new taxon to existing species and classifications. Many of these 'new' species will have been described before, so sorting out the mess will be the headache of the next generation of taxonomists. It is not surprising if funding bodies view much of what taxonomists do as poor value for money.

One of the astonishing things about

This discipline is made for the web: it is information-rich and often requires copious illustrations.



From paper to screen: is it time for taxonomy to break with tradition and unify on the Internet?

being a scientist at this particular time in history is the vast amount of information that is available, essentially free, via one's desktop computer. I can download the sequences of millions of genes, the positions of countless stars. Yet, with a few wonderful exceptions, the quantity of taxonomic information available on the web is pitiful, and what is present (typically simple lists) is of little use to non-taxonomists. But surely taxonomy is made for the web: it is an information-rich subject, often requiring copious illustrations. At present, the output of much taxonomy is expensive printed monographs, or papers in low-circulation journals available only in specialized libraries. These are not attractive 'deliverables' for major research funders.

Two models of taxonomy

The taxonomy of a group of organisms does not reside in a single publication or a single institution, but instead is an ill-defined integral of the accumulated literature on that group. The literature is bound together and cross-references itself using the venerable rules of taxonomy encapsulated in the codes. But this is not the only way to organize a taxonomy. The taxonomy of a particular group could reside in one place and be administered by a single organization. It could be self-contained and require reference to no other sources.

My main argument is that to address the problems outlined above, and for taxonomy to flourish now and in the future, it has to move from the first to the second model: from having a distributed to a unitary organization. Such a massive task could only be accomplished group by group, as resources became available. I believe a number of things would then follow. First, the only logical way to organize a unitary taxonomy and to make it widely available is on the web. The web is currently used, if used at all, as an adjunct to the distributed, printed taxonomy, but I think it should replace it. Second, the core of taxonomy is a description of each species and a means of distinguishing among them; to this core has been added the exercise of resolving their evolutionary relationships. I believe that taxonomy needs to expand to include other aspects of the species' biology, to become an information science that curates our accumulated knowledge of that species in the way a gene annotation in a genome database organizes our knowledge of a particular protein. Third, I think it is essential that the unitary taxonomy of different groups evolves from the present taxonomy. We must preserve the achievements of 250 years of distributed taxonomy, dispensing with the bad legacy of the past but retaining the good.

To illustrate how this could be done I

Box 1: Taxonomy on the web

The current codes of zoological and botanical nomenclature do not allow original descriptions to be made purely on the web, but nevertheless there is a substantial amount of taxonomy on the Internet. The Natural History Portal of the Natural History Museum in London (www.nhm.ac.uk/portal/index.html) provides an excellent entry into these resources, which include such sites as the International Plant Name Index (www.ipni.org) that covers all higher plants; the ant database (www.antbase.org) featured recently in *Nature's* News section (416, 115; 2002); and the Tree of Life project (tolweb.org/tree), a database of phylogenies.

The most common data available are catalogues of species names and lists of museum specimens, although some identification keys and other information-rich sites are becoming available.

An ambitious project led by Species 2000 (www.sp2000.org) and the Integrated Taxonomic Information System (www.itis.usda.gov) aims to catalogue the world's biota, and these sites themselves also link to the Global Biodiversity Information Facility (www.gbif.org), intended to be a general clearing house for biodiversity information.

Finally, the All Species Foundation (www.all-species.org) has set itself the goal of making an inventory of all species on Earth in the next 25 years.

shall sketch one possible way a unitary taxonomy might be achieved. I am not a professional taxonomist and am under no illusion that what follows will be the best or even a viable model, but I hope it will bring out the issues involved.

A unitary taxonomy

Introduce as a formal taxonomic procedure the 'first web revision'. This would be a revision of a major group of organisms to a standard decided on by the International Commission on Zoological Nomenclature, or the International Botanical Congress, or equivalent body (let's just call it the international committee). The revision would include a traditional description of each taxon and the location of type material. It might also include material not currently required in a formal description, for example keys and, for many groups, photographs or other illustrations. For some organisms a gene sequence might be required. It would also include a treatment of existing known synonyms to preserve contact with the older literature. This draft first web revision would be placed on the web for comments from the community, then after changes have been made in response, it would become the unitary taxonomy of the group.

What would this mean? First, from this time onwards all future work on the group need refer only to the set of species in the first web revision and then later to those in the 'nth (that is, current) web revision'. The taxonomy of the group is thus at a stroke liberated from nineteenth-century descriptions and potentially undiscovered synonyms. If I think I have discovered a new species I need only to check that it is not already in the web revision. So what happens if I describe a new species and then someone discovers that Linnaeus or someone had already described it in an overlooked work? Well, that interesting nugget of historical information can be added to the species' web page, but the name doesn't change. What happens if I want to lump, split or add species, or revise their higher classification? Then I submit a revision that is mounted on the web for refereeing and comment. If, as a result, it is accepted, it becomes incorporated into the current ($n+1$ th) web revision. At any one time there is just a single current web revision to which people refer, linked to all previous revisions (which are maintained on the web, so that in future I can easily see what was understood by species x in year y).

A major difference between this way of doing taxonomy and the *status quo* is that a unitary taxonomy needs administration: both the physical implementation on servers and networks, and the intellectual administration of the current web revision. One virtue of the present system is that if no one is interested in a group's taxonomy it can quietly slumber in the library. But the collections and

type material that underpin distributed taxonomies do require administration, which is currently undertaken by our great museums and herbaria. Nearly all these organizations are enthusiastically embracing modern web technologies. Hosting web revisions is something I see as a logical extension of their moves towards becoming, in part, modern information storehouses. It is absolutely clear, however, that they need more money in order to do this. They might also undertake the intellectual administration of the web revision — the refereeing and editing — although they would probably devolve this to committees drawn from a wider constituency (the equivalent of a journal's editorial board).

However it worked, standards would need to be set and monitored by the international committee, who would also determine which institute houses which taxonomy, and would prevent duplication of effort.

Advantages

I believe that what I have described is evolutionary rather than revolutionary in that it preserves the hard-won successes of current taxonomy while dispensing with the historical baggage. It is also evolutionary in that groups would move to the new unitary taxonomy as resources became available. It would set a series of achievable targets that could be used to spur major funding initiatives, for example the first web revision of mosquitoes, reptiles or plants (and I hope *Nature* or *Science* might celebrate these milestones as they do completed genome sequences).

I believe that major government and private research funders would consider construction and maintenance of a unitary taxonomy — universally accessible, and the foundation of all future work on the group — much more attractive to support than taxonomy as presently practised. It might also attract new sources of funding. It surely isn't impossible that a major company might sponsor the web revision of, say, the Lepidoptera (butterflies and moths); and if it wants to put its logo on the site, then why not?

The web revision would become an information hub, both through its contents and through its links to other sites. Links to molecular databases will facilitate the increasing usefulness of molecular techniques in species identification. There are already exciting web-based phylogenetic projects (see Box 1) that aim ultimately to build a phylogeny of all living organisms; clearly, one would build in reciprocal links to these sites. Today, a reference to a species in a scientific article usually gives just the scientific name and possibly the authority, but seldom refers (or gives credit) to the taxonomic revision upon which the identification is based. As increasing numbers of journals go electronic, the mention of a species can more and more easily be linked to its position in the current web revision. Were the status of the

species to change, the link would take you to the contemporary web revision and then forward to the current conception of the taxon. These links could also be used to produce a much-needed, fair 'citation count' for taxonomists. Finally, as an increasing amount of the scientific literature becomes available online through projects such as JSTOR (www.jstor.org), one can imagine links between a species description and important early papers on its taxonomy and biology, again maintaining links with the good legacy of distributed taxonomy.

Many taxonomic works are very hard for non-specialists to use, sometimes because of real difficulties in telling many species apart, but more often because of the telegraphic jargon and lack of illustration imposed on taxonomists by the expense of publication in print. The web has far fewer constraints, and provides the space needed for taxonomists to be understood. Taxonomy often pays insufficient attention to its 'end users', the ecologists, conservationists, pest managers and amateur naturalists who need or want to identify animals and plants. I hope that, overlaid on the current web revision, there would be higher-level information, the equivalent of the regional field guides and floras used by field workers. For many, this 'entry level' would be all that is required, but where needed the user could burrow deeper, right through to the primary taxonomic sources. Today, few people would seriously think about taking a computer into the field as a substitute for a field guide, but that will undoubtedly change and taxonomists should be ready.

Finally, the taxonomy should be available free (without access charges) to anyone who can log onto the Internet. This will raise the profile of taxonomy and increase the number of people who actually use the fruits of taxonomic research. Longer-term positive benefits will be for a new, young generation of naturalists, stalking their prey using digital cameras, downloading their captures into PCs, then identifying them over the web — exposing them to taxonomy as an active discipline, at the heart of modern biology.

Disadvantages

One disadvantage of a unitary taxonomy is the requirement for more administration, with its attendant costs. My assertion is that the advantages of a unitary taxonomy will prime sufficient new funds to counterbalance this, but if I'm wrong the project fails. There are also considerable technological challenges in developing the web software to support the taxonomies.

A possible criticism is that the proposal is top-down, at variance with the individualistic tradition of taxonomy. Would one clique be able to impose its view of how a group is classified? The international committee would be empowered to set standards, but rejected contributions to a group's taxono-



Harnessing the power of the web would allow all contributions to taxonomy to be collated.

my should also be stored on the web. Even if they are not incorporated in the current web revision they can at least influence future scholarship and research.

An important issue is the degree to which a treatment should be 'complete' before it is a candidate for a first web revision. Could a series of intractable species complexes requiring detailed research delay completion of a revision? The ideal solution would be to commission new taxonomic research to sort out these problems, but if this is not possible I would favour a category of 'provisional taxon', where the need for further study is clearly highlighted. After all, the heterochromatin-rich gaps in the human genome sequence did not delay the announcement of its 'completion'.

Is a web-based taxonomy as permanent as a paper-based one, and are people without computers disenfranchised, especially those in less wealthy countries? I believe the first is a non-issue; there is not (as far as I know) a paper back-up to the human genome database, and the international committee would set rigid standards for archiving and backup. Access is a much more important matter, but very many more people are at present disenfranchised by their inability to get to a specialist library, or to order a reprint, or even by being unaware that certain literature exists. The web-based taxonomy must be completely downloadable so that even continuous access to the Internet is not essential, and, if all else fails, a paper copy could be printed. It might spread the geographical distribution of taxonomic activity if some sites were hosted by developing countries with strengths in computing, such as India.

Conclusions

I find that the commonest reaction of taxonomists to these ideas is the worry that it is an attempted technological fix that distracts attention from what they (and I) perceive to be the overwhelmingly critical issue — the lack of people and resources devoted to descriptive taxonomy. The counter-argument is that the technological fix is not an end in itself; it is the means of making grassroots taxonomy more accessible and useful,

and thus attracting people and funds into the field. But is such a root-and-branch change in the culture of taxonomy really needed? Although there is near-universal agreement about the current depressed state of descriptive taxonomy, wouldn't more funding alone solve the problem?

I think not: indeed, descriptive taxonomy might disappear completely for 'difficult' groups such as many insects and nematodes. Just as Moore's law says that microprocessor power doubles every 18 months, there must be a parallel law that says DNA sequencing power increases geometrically. In 10 or 20 years' time it will be simpler to take an individual organism and get enough sequence data to assign it to a 'sequence cluster' (equivalent to species) than to key it down using traditional methods, let alone describe it as new. Just as bacterial taxonomy is now nearly all sequence-based, a new way of classifying insects, nematodes and perhaps even many plants and fish might evolve that is totally divorced from current taxonomy — a point also made forcibly by Robert May, president of Britain's Royal Society.

Would the death of large swathes of present-day systematics matter? Yes it would, because we would be throwing away so much of what we have learned in the past 250 years about the planet's biota, a lot of which we would then have to relearn. But unless taxonomy is unitary, web-based and able to accommodate these radical new ways of doing biology, I fear it will be sidelined.

The rigidity built into the current rules and codes of taxonomy — which include prohibition of purely electronic description — is part of their success, and changes should not be made lightly. But I suspect these rules are now a brake on progress, imprisoning the subject in outdated methodologies, and rendering it difficult or impossible to attract the major funds needed to reverse its slow decline. Surely it is time to experiment — time for the international taxonomic community to come together and countenance a unitary web revision of one or a few major groups of organisms (and to work out exactly how a unitary taxonomy should operate). This venture must be sanctioned and supported by the existing international committees, or no serious taxonomist will waste his or her time on it; no institution will administer it; and no agency will fund it. If successful, it will change how taxonomy is done for ever; if it fails it would not be difficult to revert to the *status quo ante*. There is everything to gain and little to lose. ■

H. Charles J. Godfray is at the NERC Centre for Population Biology, Department of Biological Sciences, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK.

Acknowledgements

I am grateful to the many taxonomists and other biologists who have debated these issues with me.

Emi's fate, our fate

A stark warning that overpopulation is threatening global biodiversity.

The Future of Life

by Edward O. Wilson

Little, Brown: 2002. 244 pp. £18.99 (UK);

Alfred A. Knopf: 2002. 256 pp. \$22 (US)

Paul R. Ehrlich

Edward O. Wilson is a member of an important but very rare species: the world-class scientist who is also a great writer. The prologue to *The Future of Life*, written as a letter to Henry David Thoreau, the “founding saint of the conservation movement”, illustrates his graceful style beautifully. From contemplating Walden Pond in Massachusetts, Wilson moves on to consider the state of global biodiversity. He accurately and passionately tells the story of the disappearance of many of the only living beings we know of in the Universe — key components of humanity's natural capital.

The Future of Life outlines for the non-scientist the causes of the extinction crisis, explains the stake that humans have in it, and analyses solutions — some of which are already being put into practice. Even biologists who are familiar with much of the book's content will find it a wonderful read, full of interesting details. Wilson has a splendid feel for organisms, whether he is telling the story of Emi, one of the remaining handful of Sumatran rhinos, or describing the intricate biological community that could be found in a rotting log in a vacant lot ‘micro-reserve’.

I was especially pleased to see a clear statement that human population growth plays a key role in the loss of biodiversity and general degradation of the environment. A misguided political correctness has prevented many scientists and environmentalists from stating the obvious — that overpopulation of *Homo sapiens* is a critical element both in causing environmental problems and in exacerbating their consequences. The more people there are, the more greenhouse gas is emitted into the atmosphere, all else being equal. The more greenhouse gas, the greater the chances of rapid climatic change, which could stress agricultural systems. The more people there are, the harder it will be for those stressed systems to feed everyone.

Wilson is direct. Speaking of the United States, “whose citizens are working at a furious pace to overpopulate and exhaust their own land and water from sea to shining sea”, he paints a picture completely at odds with the uninformed view of most Americans.

He also briefly addresses a less-recognized companion problem to overpopulation: overconsumption. Writing to Thoreau, he



Paradise lost? The Earth's rich biological heritage is at risk from human population growth.

says: “We are inside a bottleneck of overpopulation and wasteful consumption” (my emphasis). The consumption multiplier (which in a sense makes the United States the most overpopulated nation) is not a popular topic in consumer societies. Appeals for conservation and efficiency often fall on deaf ears, as does talk of moving “down the food chain to a more vegetarian diet”. But such talk, and discussions of consumption control in general, are badly needed. As a chocoholic and pilot, I know the challenges personally. World-class economists and ecologists are beginning to tackle the difficult problems of defining overconsumption

and dealing with it; Wilson's book could help to move the issue towards the political agenda. We already know what needs to be done to promote gradual population reduction. Maybe there will be some “consumption condoms” in our future.

It is incumbent on reviewers to say what they didn't like about a book, but that is a tough assignment in this case. Wilson and I differ mostly in matters of emphasis in areas that are far from settled. I am less sanguine than he is about the theoretical ability of the Earth in 2100 to support 10 billion people at a “decent standard of living”, and believe that the problem of population extinctions

(which degrade critical ecosystem services) deserves more attention.

On the positive side, I think that the development of countryside biogeography as a framework for enhancing the preservation of biodiversity in human-dominated landscapes deserves attention alongside prospects for establishing large-scale reserves, which Wilson discusses very thoroughly. On a separate issue, I'm more sceptical about heritability estimates churned out by behavioural geneticists (often based on badly analysed twin studies) for such attributes as proneness to agoraphobia and fear of snakes. But these are trivial matters compared to the magisterial sweep of *The Future of Life*, and I find myself in total agreement with its major points.

Wilson was recently attacked viciously in the pages of *The Economist*. He was critical of Bjørn Lomborg's anti-environmental book *The Skeptical Environmentalist*, which the magazine and Cambridge University Press have been heavily promoting. In my view, Wilson had accurately pointed out that busy scientists were having to waste a huge amount of time replying to the book's distortions. *The Future of Life*, by coincidence, is Wilson's perfect response. It clearly lays out the reasons for his deep concern for the human future (shared by the vast majority of his colleagues) and why he thinks that scientists and society have no time to waste. It also reveals him to be a thoughtful, caring, life-loving human being. ■

Paul R. Ehrlich is in the Department of Biological Sciences, Stanford University, Stanford, California 94305, USA.

From *E. coli* to elephants

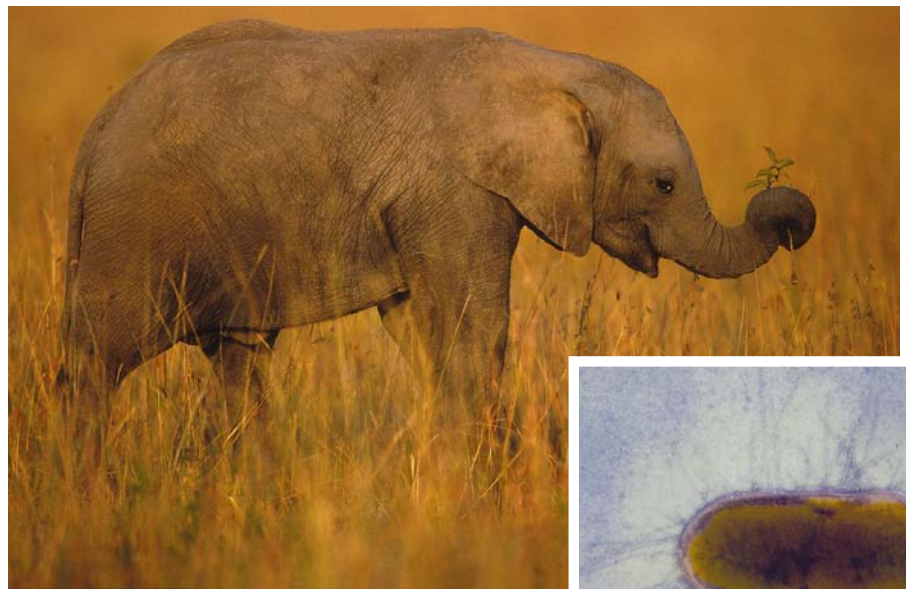
Genes and Signals

by Mark Ptashne and Alexander Gann
Cold Spring Harbor Laboratory Press: 2002.
208 pp. \$59, £43 (hbk); \$39, £28 (pbk)

Kevin Struhl

The regulation of gene expression is a fundamental aspect of biological phenomena such as the response to environmental conditions, the development of multicellular organisms, morphology and disease. Gene regulatory patterns are extraordinarily diverse and complex, yet the regulation of each gene is precise with respect to when and how much expression occurs. Gene regulation is also remarkably flexible, both to rapidly alter the constellation of genes expressed in response to new conditions, and to accommodate evolutionary demands. At most, a few thousand proteins account for the complexity and precision of gene regulation in humans. How is this accomplished?

Molecular studies of gene regulation were



The same, only more so: gene regulation is similar for organisms ranging from the elephant to the bacterium *Escherichia coli*.

pioneered by François Jacob and Jacques Monod in the early 1950s. By the mid-1960s, three basic types of specific DNA sequence that determine the level of expression under particular physiological conditions were defined in the bacterium *Escherichia coli*. Such regulatory DNA sequences turn out to be specific binding sites for RNA polymerase, repressor proteins and activator proteins. Regulation of an individual gene is determined by the quality of its polymerase binding site, the particular activator and/or repressor proteins that bind in the vicinity of RNA polymerase, and the physiological conditions that modulate the function of the activators and/or repressors.

Monod once wrote that "anything that is true of *E. coli* must be true of elephants, except more so". In a lucid and provocative book, Mark Ptashne, a leading figure in the field for nearly 40 years, and Alexander Gann argue for a unifying principle of gene regulation that centres on the concept of regulated recruitment by means of adhesive interactions between proteins. They go on to argue that such regulated recruitment is a general strategy used by many other biological mechanisms involving enzyme specificity, regulatory precision and evolutionary flexibility.

Using a few well-chosen examples, Ptashne and Gann first describe three distinct mechanisms of transcriptional activation in bacteria. In one mechanism, DNA-binding activator proteins stimulate gene expression by recruiting RNA polymerase to the promoter sequences that lie just upstream of the gene. Recruitment is mediated by short 'adhesive' surfaces between the activator and polymerase, and the adhesive properties *per se* are sufficient for activation. In a second mechanism, the activator induces a

conformational change in an inactive polymerase already bound at the promoter, thereby stimulating transcription. And in a third mechanism, the activator induces a conformational change in the promoter, effectively changing it from an inactive to an active form. This section of the book presents the key experiments and arguments for these mechanisms in a manner that is exceptionally lucid and beautifully illustrated. It is understandable to the non-expert, for whom it was intended, and is a 'must read' for anyone interested in gene regulation.

Armed with these lessons from bacteria, Ptashne and Gann consider yeast, a single-celled eukaryote, and conclude that activation occurs by regulated recruitment of the transcription machinery (which contains more than 50 proteins and so is much more complex than bacterial polymerases). Again, the authors use the device of a well-chosen example for clarity, the brief is convincingly argued, and the end result is illuminating to both the expert and the novice. The emphasis on regulated recruitment is important for the overall theme of the book, and it is certainly true that this mechanism predominates in yeast cells.

However, in emphasizing the fundamental similarities between bacteria and eukaryotes, Ptashne and Gann have made an unconventional choice in classifying chromatin-modifying activities as part of the transcription machinery. Chromatin and chromatin-modifying enzymes affect all eukaryotic processes involving DNA, and are typically considered as part of the DNA template, rather than the transcription machinery. So although activators and repressors use adhesive surfaces for regu-



lated recruitment of chromatin-modifying activities, such recruitment does not directly affect transcription and indeed is analogous to (although mechanistically distinct from) the bacterial mechanism in which activators modify promoter structure. In eukaryotes, the basic chromatin structure renders core promoters inherently inactive in the absence of an activator, whereas bacterial promoters are generally accessible to the polymerase. In my view, chromatin fundamentally affects the logic of gene regulation in eukaryotes, but fundamentalism is in the belief of the beholder.

The only disappointing part of the book is the brief section on higher eukaryotes. Unlike the rest of this book, and unlike Ptashne's previous influential monograph *A Genetic Switch*, this section covers many different phenomena (all very interesting and important) in a rather sketchy fashion. This subject calls for another book, although it is probably premature to write one at the level to which we have become, and wish to remain, accustomed.

In the grand scheme, the principles of regulated recruitment through weak, adhesive interactions between proteins are applied to other examples of enzyme specificity and regulation (such as splicing, proteolysis and signal transduction), where diversity, precision, and evolutionary flexibility are paramount. Some may consider this section to be a statement of the obvious, namely that protein-protein interactions are important in biology. However, I agree with Ptashne and Gann that this concept is fundamental to understanding specificity in biology, and is, in historical context, a revolutionary idea. Until recently, enzyme specificity and protein function were thought of in terms of precise active sites with near-unique substrate recognition. Regulation in this view occurs by allostery, a mechanism in which a signalling molecule alters the enzyme in a structurally precise manner.

In this context, the emerging picture that a great deal of biological specificity is mediated by simple adhesive interactions involving limited and modular protein surfaces is neither obvious nor intuitive. But the concept has the undeniable virtue of explaining the apparently contrary notions of precision and flexibility. In addition to the exposition of this major theme, the final section of the book and the afterword are full of interesting insights.

In *Genes and Signals*, Ptashne and Gann have written a unique book that is driven by ideas and broad concepts, yet is based on solid information. It is accessible to undergraduates with some knowledge of biology, yet it is also valuable to experts in the field. I highly recommend it. ■

Kevin Struhl is in the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA.

Science in culture

Seeing stains

Mary Osborn's immunofluorescence images of cellular structures.

Martin Kemp

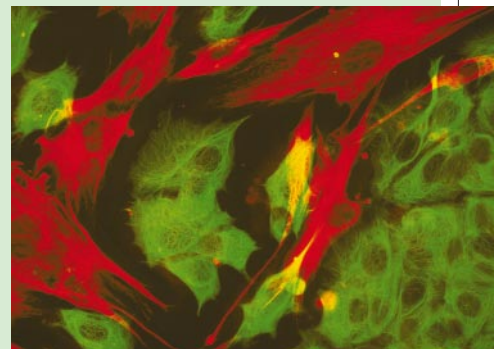
Stains (and those who have pioneered their use) are the unsung heroes of microscopy — well known to microbiologists, certainly, but not generally enjoying a high public profile. New instruments for seeing ever smaller details seem to present more eye-catching examples of scientific advance.

The story began with the progress from the primitive microscopes of the seventeenth century, in the hands of such pioneers of discriminating seeing as Anthony von Leeuwenhoek and Robert Hooke, and the gradual refining of optical resolution to its theoretical limits. Then came the non-optical revelations of electron microscopes, as developed by Vladimir Zworykhin and others, and the molecular marvels disclosed by the scanning tunnelling microscope devised by Gerd Binnig and Heinrich Rohrer. Yet without selective staining and other marking techniques we would not be able adequately to differentiate key components in the tiny structures.

One of the most elegant and widely applicable techniques, immunofluorescence, has, as its name suggests, ingeniously adopted techniques from immunology. Albert Coons used it to reveal pneumococci in infected mouse tissues in 1942. But its power in visualizing intracellular structures was only realized in 1974 when Elias Lazarides and Klaus Weber (then at Cold Spring Harbor Laboratory in New York state) used it to reveal microfilament bundles in cells. The basic technique works as follows.

Given a protein (antigen) that is of interest, an antibody is made. Once the cell has been made permeable, the antibody invades and binds itself to the antigen. The unbound antibody is then washed away. A second antibody carrying a fluorescent marker (such as green fluorescein or red rhodamine) recognises the first, and the process of washing away the excess is repeated. It is rather like doubling the process in conventional photographic printing with 'hypo', in which the excess developer must be fully washed out if the image is to become clearly defined. The final step of immunofluorescence is to view the marked protein using a microscope with a light source and filter set.

A pioneer of the technique, Mary Osborn of the Max Planck Institute of Biophysical Chemistry in Göttingen, Germany, is the European winner of the 2002 L'Oréal/UNESCO Prize for Women in Science, awarded for international scientific excellence. She has also been in the vanguard of promoting the status of women scientists. Her work demonstrates that the results of immunofluorescence are both scientifically potent and visually beguiling to a high degree. Osborn, working at times together with Weber, has used immunofluorescence to



Colour coding: an immunofluorescence micrograph of epithelial and fibroblastic cells.

disclose microtubules and intermediate filaments as functional components in cells.

The story of the microtubules vividly shows how optical and electron microscopy need to work hand in hand, and how the greater magnification of the latter does not necessarily deliver fully coherent results when visualizing the structural continuities of forms that extend across the cell. In the late 1970s there was an acrimonious debate about the length and number of microtubules. Weber and Osborn were even accused of "painting white lines" on their images. Only when the same cell was visualized by the Göttingen group, both with immunofluorescence staining for tubulin and as a whole mount in electron microscopy, did the evidence of extended microtubules speak unequivocally for itself.

The intermediate filaments have proved to be particularly useful in routine pathology and cytology, acting as highly effective markers of cell type. As Osborn explains, "human tumours retain the intermediate filament typical of their origin. Intermediate-filament antibodies can be of particular use in certain differential diagnoses when the pathologist or cytologist is unsure of the diagnosis by conventional staining."

The micrograph shown here was made by Osborn in 1987, and shows a mixture of epithelial and fibroblastic cells growing in culture stained with antibodies against two intermediate filament proteins, keratin and vimentin. The keratin antibody decorates the intermediate filaments present in the epithelial MCF7 cell line (in green) whereas the vimentin antibody decorates the filaments in the fibroblastic HS27 cell line (in red), thus distinguishing the two cell types.

Alongside the technical achievements lies a sheer love of looking at the perpetually varied topography of stained cells. As Osborn says, "I can still stare down the microscope for hours. Not only because they are beautiful but also because every cell shows subtle differences in the arrangement and distribution of the three filament systems." ■

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

Darwin's motors

George Oster

In 1827, Robert Brown famously observed under his microscope pollen grains dancing as if alive. At first he thought he might be observing the “elementary molecules of organic bodies” — the life force itself. And, in a sense, he was. The pollen grains were not ‘alive’, but were being driven by the thermal motions of the surrounding fluid molecules.

A complete explanation of Brown's observations came with Einstein's 1905 theory of ‘brownian motion’ and its experimental confirmation by Jean Baptiste Perrin, who won the 1926 Nobel Prize for his work (Einstein's theory did not — what were they thinking?).

Brown was right about the movements he observed being the ‘life force’ because the chemical reactions on which life depends are driven by brownian motion. Enzymes must explore millions of configurations per second to seize on the correct one that allows a reaction to take place. Thermal fluctuations at the molecular scale are almost unimaginably tumultuous, and are easily up to the task.

Motor enzymes use energy stored in chemical bonds to generate directed forces for an amazing variety of tasks. Despite their variety, they all operate on the same principle: they trap brownian fluctuations, albeit in different ways. One mechanism involves biasing, or rectifying, the otherwise random brownian movements by using short-range attractive forces to trap favourable fluctuations. For example, a polymer can push a load ahead of it

as it polymerizes. Here the brownian motion of the load (and of the polymer itself if it is flexible) eventually opens up a gap large enough for a monomer to insert itself between the load and the polymer tip. Thus the load is prevented from diffusing backwards once the monomer is in place — it is driven forwards in monomer-sized steps, using the energy that binds the monomers together to ratchet the fluctuations of the filament and the load.

This ‘brownian ratchet’ mechanism is how many intracellular parasites propel themselves through their host's cytoplasm. A similar process moves unfolded proteins through the membranes of cell organelles: ‘trapping’ proteins bind to the portion of the target protein diffusing through a transmembrane pore, preventing it from going backwards again.

The speed of this kind of ratchet varies inversely with the size of the load, so large loads tend to move slowly. Moreover, the motor can be used only once, and is then disassembled into its subunits and reassembled anew. But a protein's catalytic site is flexible so that it can respond to smaller, and therefore much more frequent, brownian fluctuations, rather than waiting for the load to diffuse over a large distance. These smaller ratchet steps enable motor proteins to form direct, elastic couplings with their load and to drive it with a power stroke using the energy acquired by binding to a substrate protein.

This binding is progressive: the protein wraps around its substrate, each step driven by ångström-sized brownian fluctuations that bring the protein and substrate atoms within range of intermolecular attractions that trap the fluctuation. (The wrapping of the catalytic site around the substrate is stochastic, so fluctuations in the opposite direction can unwrap as well, but the intermolecular attractions favour wrapping.) As the intermolecular bonds between enzyme and substrate ‘zip up’, they induce a strain in the binding site that is used to lever the load over distances that are comparable to the size of the protein itself. The biochemist Daniel Koshland first proposed the idea of an ‘induced fit’ between an enzyme and its substrate, although he was not thinking specifically about motor proteins.

Enzymes that use this ‘small fluctuation’ mechanism have the further advantage that they can operate in a continuous cycle by using a ‘fuel’ molecule — usually ATP in cells — as a binding partner. After binding to the motor and driving the power stroke, this nucleotide can be cleaved into two. Each piece can then be knocked out of the binding site by brownian fluctuations, freeing the site to bind to another ATP molecule. Thus brownian motion drives both the power and exhaust strokes.

Many motor enzymes combine both of

Brownian ratchets

The molecular motors on which life depends are driven by brownian motion.

these mechanisms to generate a directed force. Some resemble two ‘legs’ that step along a polymer track between equally spaced binding sites. The ‘front’ leg of the enzyme is driven forwards, and then ‘hunts’ by diffusion for the next binding site along the track. When the front leg is close enough to the next site, the excursion is trapped and a ratcheting of brownian motion has taken place. These ‘walking motors’ use two binding partners: attachment to the track sites generates a power stroke, whereas cycling of ATP provides the energy to break free for another step. The physiologist Andrew Huxley first proposed a mechanism of this sort for myosin, the motor protein that drives muscle contraction.

One of the most remarkable motor enzymes is F_1F_0 ATPase, possibly life's most abundant protein, which catalyses the production of ATP. (Every day, we produce — and consume — about half our body weight in ATP!) This enzyme consists of two rotary motors attached to a common shaft. The F_1 motor generates a power stroke using ATP as its fuel; the F_0 motor is almost a pure brownian ratchet that uses the binding and release of protons flowing through it to rectify its rotational diffusion. More than any other, this protein has illuminated our knowledge of the miniature motors that form the true basis of Brown's ‘molecules of life’.

In a broader sense, the idea of generating order by ‘selecting’ from random variations is hardly new — it is the fundamental idea of Darwin's theory of natural selection. In the context of motor proteins, the ‘order’ created is a directional force, and the agents of selection are intermolecular attractions. Hence the idea of a brownian ratchet keeps popping up in new contexts, providing a fertile stimulus to our thinking in disparate fields. Indeed, as the philosopher Daniel Dennett has said — and I agree — Darwin may have had the best idea that anyone ever had. Think about it. ■

George Oster is in the Department of Molecular and Cellular Biology, University of California, Berkeley, California 94720-3112, USA.

FURTHER READING

Huxley, A. F. *Curr. Biol.* **8**, R485–R488 (1998).
Boyer, P. *Angew. Chem. Int. Edn* **37**, 2296–2307 (1998).
Huxley, A. F. & Simmons, R. M. (eds) *Phil. Trans. Biol. Sci.* **355**, 413–543 (2000).
Walker, J. E. (ed.) *Biochim. Biophys. Acta* **1458**, 231–514 (2000).



Molecular ratchets mirror mechanical ones.

Something new under the sea

Yan Boucher and W. Ford Doolittle

The discovery, in an undersea hot vent, of an organism that does not fit into any previously defined category of life marks the creation of yet another group within the mysterious Archaea.

An exciting new creature has been discovered under the sea off Iceland, and it is described by Huber *et al.* on page 63 of this issue¹. Although invisible to the naked eye, it is as worthy of our notice as any coelacanth or other macroscopic 'living fossil', for three reasons. First, its discovery means that current methods are still inadequate to reveal life's true diversity. Second, it is either very primitive or extremely modified by evolution. Finally, this newly identified entity may represent a whole new group within the Archaea, the most recently discovered and still most mysterious of life's three domains (the others being the Eukaryota and Bacteria).

Much of life's diversity is microbial, but most microbes cannot be grown in culture. How, then, can we discover anything about such organisms? In particular, by what methods can we establish their relatedness to those few microbes that we are able to cultivate and study in the laboratory? The answer is the polymerase chain reaction (PCR). This technique can be used to amplify — make many copies of — DNA isolated directly from environmental samples; the DNA can then be sequenced. Comparing sequences with each other and with those stockpiled in databases gives us a measure of true microbial diversity².

Genes that code for the RNA component of ribosomal small subunits (SSU rRNA) are the best targets for amplification, sequencing and comparison. This is because ribosomes — cellular complexes of rRNA and protein that are the site of translation of the DNA code — are essential for the survival of any organism. So all organisms carry at least one rRNA gene, and rRNA gene sequences have changed relatively little during the 3–4-billion year history of life on Earth. They are accepted by microbiologists as the common currency of microbial taxonomy. Indeed, it was rRNA sequence analysis that led to the initial discovery of the Archaea, a group now known on many grounds (lipid biochemistry, and the fundamental machineries of DNA replication, transcription and translation of RNA, for instance) as being distinct from Bacteria.

Using PCR 'primers' to target specific regions of bacterial, archaeal or eukaryotic SSU rRNA genes makes modern-day microbe hunting even more selective and efficient. In this way, the number of known

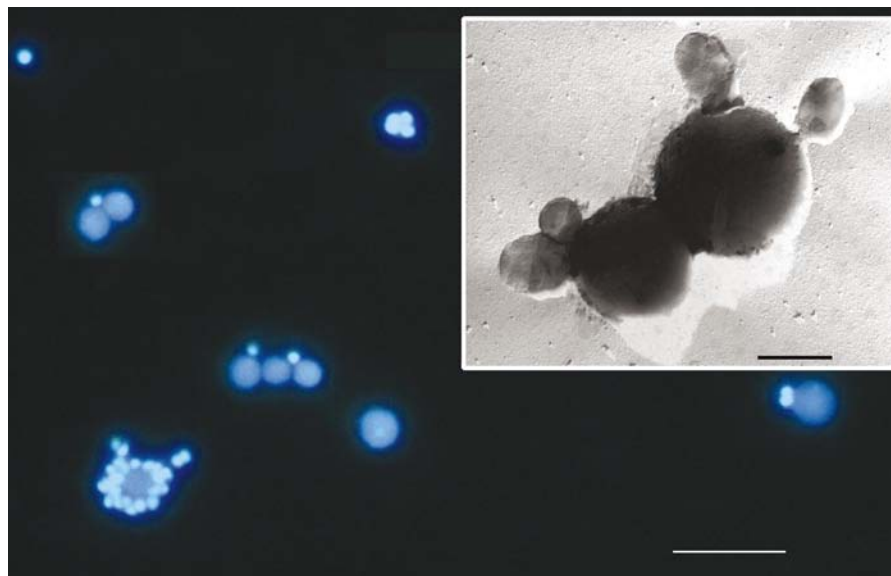


Figure 1 The little picture for microbe hunters. The main image is a fluorescence micrograph, taken after DNA-specific staining, which shows the newly discovered *Nanoarchaeum equitans* attached to its host, *Ignicoccus*. The two organisms are shown again in the inset, an electron micrograph. Scale bars are 5 μm and 1 μm , respectively. The investigations of ribosomal RNA gene sequences carried out by Huber *et al.*¹ show *N. equitans* to be a decidedly weird member of the Archaea.

major divisions within life's two prokaryotic domains, Bacteria and Archaea, has doubled^{3,4}. And just last year, two groups reported finding totally unknown, tiny eukaryotes among the 'picoplankton' of the Pacific and along the Antarctic polar front^{5,6}. (Prokaryotes are organisms that have no intracellular organelles or nucleus; eukaryotes possess organelles and a nucleus for their DNA.)

Satisfying as this has been, there is a lingering suspicion — or romantic hope, depending on your disposition — that there might be even weirder organisms hiding out there. Such organisms could be especially primitive or wildly divergent creatures that have drifted so far out of the biological mainstream that the usual PCR primers don't recognize or amplify their SSU rRNA genes. Huber *et al.*¹ provide a stunning example of just such a weird bug, something they call *Nanoarchaeum equitans* (Fig. 1).

Nanoarchaeum was discovered as small (400-nm) spherical structures attached to the cells of a much larger organism, a new species of the archaeal genus *Ignicoccus* taken from rock and gravel near a hot submarine vent north of Iceland⁷. Specific *in situ* staining procedures showed that these tiny cells contain DNA. Although several SSU rRNA-

specific primers were used, only one sort of SSU rRNA could be amplified by PCR from the mixed culture. However, two kinds of SSU rRNA genes could be detected by Southern blot hybridization, a method that is less specific than PCR and which can be used to detect (but not amplify) most rRNA genes, no matter how unusual. One type of SSU rRNA gene comes from *Ignicoccus*; the other could be shown to be localized to the tiny *Nanoarchaeum* cells. Sequencing suggests that this second gene is a true, functional SSU rRNA gene, the product of which could form a normal RNA secondary structure. And although the gene's sequence is in general closer to those of archaeal than bacterial SSU rRNA genes, it is very divergent; only specially redesigned PCR primers could detect it.

Cells of *Nanoarchaeum* can be purified from mixed cultures, and the total length of the DNA molecules they contain is roughly 500 kilobases. This is possibly the smallest-known prokaryotic cellular genome. It lies in the range for species of *Mycoplasma*, bacterial parasites that hold the current record as minimalists among the known organisms that can grow outside other cells. So far, *Nanoarchaeum* has not been grown

M. HOHN & K. STETTER/R. RACHEL & K. STETTER

alone, even on *Ignicoccus* cell extracts or in the same vessel as live *Ignicoccus* (but separated from them by permeable membranes).

Where does *Nanoarchaeum* belong, taxonomically or phylogenetically — in terms of its evolutionary relationships? Its SSU rRNA sequence puts it in the archaeal, not the bacterial, domain. Archaea comprise the Crenarchaeota (such as *Sulfolobus*), the Euryarchaeota (*Halobacterium* and *Methanococcus*, for instance) and several possibly more deeply branching lineages (including 'Korarchaeota', so far known only from environmental DNA samples). *Nanoarchaeum*'s SSU rRNA is very different in sequence, however, and it is not clear that it can be placed within any known archaeal group. Huber *et al.*¹ plump for an early branching, before the diversification of the rest of the Archaea, and also speculate that *Nanoarchaeum* might retain many primitive features. But equally, *Nanoarchaeum* could be a highly derived descendant of some already known group, off on an evolutionary tangent of its own.

Experiences in eukaryotic phylogeny may provide an insight here. Members of at least one eukaryotic group (microsporidia), whose highly divergent SSU rRNAs were initially taken as evidence for primitiveness and early evolutionary divergence, are now widely accepted as degenerate derivatives of fungi⁸. Their divergent SSU rRNAs and drastically reduced genome sizes are regarded as consequences of the extremely parasitic lifestyle that is characteristic of the group. Perhaps this is also the case for *Nanoarchaeum*. One way to tell, as Huber *et al.*¹ suggest, would be to use PCR primers based on the *Nanoarchaeum* SSU rRNA to track down its relatives, in the sea or elsewhere. If free-living relatives were found, with more modestly divergent SSU rRNAs, then the whole group might be placed at some non-basal position in the archaeal phylogenetic tree. And if that were the case, it would mean that *Nanoarchaeum*'s small genome is reduced — degenerate or streamlined — rather than primitive, as Huber *et al.* suggest. Whatever the case, this is a genome begging to be sequenced.

The name chosen for the larger group to which *Nanoarchaeum* must belong is the Nanoarchaeota. That name seems quite appropriate, given the small size and minimalist genome, and there can be no question about the authenticity of this finding. Unfortunately, the name invites comparison with 'nanobacteria', a much more problematic group. It is unclear that nanobacteria, which are often much smaller entities (down to 50 nm in diameter), could house the minimal cellular machinery necessary for life, or that they even contain DNA⁹. *Nanoarchaeum* is nowhere near as tiny — and wherever it belongs taxonomically, it looks very real. ■

Yan Boucher and W. Ford Doolittle are in the

Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada.

e-mail: ford@is.dal.ca

1. Huber, H. *et al.* *Nature* **417**, 63–67 (2002).
2. Pace, N. R. *Science* **276**, 734–740 (1997).
3. Hugenholtz, P., Goebel, B. M. & Pace, N. R. *J. Bacteriol.* **180**, 4765–4774 (1998).
4. Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R.

Proc. Natl Acad. Sci. USA **93**, 9188–9193 (1996).

5. Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D. *Nature* **409**, 603–607 (2001).
6. Moon-van der Staay, S. Y., De Wachter, R. & Vaulot, D. *Nature* **409**, 607–610 (2001).
7. Huber, H. *et al.* *Int. J. Syst. Evol. Microbiol.* **50**, 2093–2100 (2000).
8. Keeling, P. J., Luker, M. A. & Palmer, J. D. *Mol. Biol. Evol.* **17**, 23–31 (2000).
9. Abbott, A. *Nature* **401**, 105 (1999).

Optoelectronics

Liquid crystal painting

Peter Raynes

Despite the ubiquity of liquid crystal displays, their glass 'sandwich' construction has remained unchanged. But a promising new technique could soon allow them to be mounted on almost any substrate.

Over the past 25 years, liquid crystal displays (LCDs) have found a wide range of consumer and commercial applications. From simple monochrome displays for watches and calculators, through medium-resolution displays for mobile phones, to high-resolution colour displays used in computer monitors and flat-screen televisions, they have become the principal display device for flat-panel applications. In 2000, the LCD industry was worth US\$20 billion, and almost two billion displays were manufactured.

Production of LCDs has become a sophisticated, automated, high-yield and cost-effective process, yet the underlying construction of the active, light-modulating part has remained relatively unchanged. In this issue, Roel Penterman and colleagues introduce a promising new approach to the process (*Nature* **417**, 55–58; 2002). There are still hurdles to overcome before it can be applied on an industrial scale. But it could break the existing mould for LCD construction and open up the possibility of 'painting' a display onto almost any substrate.

An LCD produced by existing techniques consists of a glass 'sandwich' formed by two glass substrates, each with conducting and alignment coatings, and a 5- μm -thick layer of liquid-crystal material such as the ubiquitous mixture E7, which is based on pentyl cyanobiphenyl and its homologues (see the authors' Table 1 on page 55). This thin layer of liquid crystal is aligned in a particular way by the substrate coatings — for example, as the name suggests, in the 'twisted nematic' device, it is twisted. Application of an electric field across the layer of liquid crystal changes the alignment of the liquid crystal and produces a switching effect on polarized light. Some progress has been made towards replacing the glass substrates with plastic ones that are lighter, more flexible and more mechanically robust. There has even been progress in the fabrication of LCDs by laminating two plastic substrates together in a

roll-to-roll process. But all of these new methods still produce the same basic sandwich construction.

Using their new technique, Penterman *et al.* effectively paint LCDs onto a single substrate. They call it 'photo-enforced stratification', or PES. The process begins with a mixture of liquid crystal, a polymer-forming monomer, an ultraviolet-absorbing dye and a photo-initiator. This mixture is then painted onto the substrate, using standard techniques such as die coating or doctor blading. Doctor blading uses a device similar to a car's windscreen wiper to remove surplus solution and leave a thin layer on the substrate.

Two exposures to ultraviolet light are then carried out. The first is with high-intensity, 400-nm light (not absorbed by the ultraviolet dye) which, with the use of a mask, creates a series of polymer walls that mechanically stabilize the structure. The second exposure uses a uniform, low-intensity, 340-nm light (which is absorbed by the dye) to form a 10- μm -thick polymer lid on top of the liquid crystal layer. The result is the formation of polymer boxes, containing the liquid crystal, that are 500 μm by 500 μm square and 10 μm high, separated by walls about 100 μm wide. All liquid crystal alignment layers and the electrodes are on the original substrate.

Penterman *et al.* present convincing evidence that they have successfully fabricated and operated a prototype low-resolution PES LCD on a glass substrate. But they also admit that much remains to be done before their technique can become a viable alternative to current LCD-production techniques.

First, the original substrate used for the PES process must contain all the electrodes needed to drive the LCD. This in itself is not a problem, but it does restrict how the liquid crystal can be used. The necessary electro-optic effect results from electric fields within the plane of the liquid crystal layer, as opposed to across it. This means that the

common switching effects (twisted nematic, supertwisted nematic and vertically aligned nematic) are not appropriate. Moreover, it is not clear that satisfactory alignment across the entire liquid crystal layer can be generated from a single aligned substrate. The authors also hint that controlling the thickness of the liquid crystal layer may be an issue; in normal LCDs, the thickness of this layer is controlled to a tolerance of 1%.

Most of all, if PES is to be adopted as a production technique, the challenge of device lifetime must also be overcome. One of the main reasons that plastic substrates have not yet replaced glass in the conventional LCD sandwich is their reduced lifetime, owing to the permeability of plastics to atmospheric contaminants such as water and oxygen. This would be particularly important if active devices such as thin-film transistors are used on the substrate. Thin-film transistors are used to drive most high-information-content LCDs and they require liquid crystal of very low conductivity. If the liquid crystal is contaminated it becomes too conducting.

However, any new technique has problems to be solved and challenges to be met before it can be widely adopted. The history of LCDs shows that industry has been highly successful in overcoming such challenges. The idea of being able to paint an LCD onto a substrate of arbitrary size and shape is so attractive that one suspects that the underlying problems will be solved. We can look forward to the day when we will be able to put displays on almost anything. ■

Peter Raynes is in the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK.

e-mail: peter.raynes@eng.ox.ac.uk

Neurobiology

The amazing astrocyte

Clive N. Svendsen

Two regions of the brain in adult mammals contain stem cells that can generate new neurons. It seems that astrocytes — cells once viewed merely as padding in the brain — can stimulate the neuron-generating process.

The cells that make up our brains come in two main flavours: neurons and glia. Neurons conduct action potentials and have traditionally held centre stage in neurobiological research. By contrast, glia have often been seen as the poor relatives. With their name originating from the Greek word for glue, they were for many years viewed simply as the brain's packing material, holding neurons in place. But this view is changing rapidly. One dominant type of glial cell is the astrocyte, and studies of its functions continue to take intriguing twists and turns. As well as providing structural support for nerve cells, astrocytes are now known to modulate the environment around neurons, release a range of neuronal growth factors and help to maintain the cellular barrier between blood and brain. They may also control neuronal life more directly by regulating the production of synapses, the brain's chemical junctions¹. On page 39 of this issue Song *et al.*² add to this impressive list, showing that astrocytes can also instruct unspecialized cells to become neurons — a process termed neurogenesis. It seems that, in the cellular hierarchy of the nervous system, astrocytes are upwardly mobile.

At first glance, the idea that astrocytes control neurogenesis appears very unlikely

— during brain development, most neurons are created much earlier than these glial cells. But matters are different in adult brains, which contain neural 'stem cells' that constantly produce new neurons in the presence of existing astrocytes. Neural stem cells, by definition, must renew themselves and produce both neuronal and glial lineages (Fig. 1, overleaf). In adult brains, this process is controlled by a combination of the stem cells' intrinsic properties and their local environment, so the exact location of adult neural stem cells is a topic of intense investigation.

In this context, the two best-studied brain regions are the hippocampus and the areas immediately surrounding the walls of the fluid-filled ventricles found in the centre of the brain and spinal cord. The 'subventricular zone' in forebrain ventricles is probably the most dynamic of the brain's neurogenic regions, daily producing thousands of new neurons that migrate into the olfactory bulb. The hippocampus likewise produces new neurons. But in most brain regions and within the spinal cord in adults, neurogenesis is very limited. If it could be enhanced and better controlled in neurogenic regions, or triggered elsewhere, it might prove possible to repair brain tissue that has been damaged through trauma

Systematics

Old insects in new order

It's a once-in-a-lifetime event. The creatures pictured here are members of the first new order of insects to be discovered since before the First World War. The Mantophasmatodea, as they have been christened, take their place alongside the other 30-odd insect orders such as beetles, flies and termites.

Writing in *Science* (online 19 April 2002; DOI 10.1126/science.1069397), Klaus Klass and colleagues have described the first two species of the order from museum specimens. Collected in Tanzania and Namibia in the early years of the last century, the specimens had languished unidentified in Lund, Sweden, and Berlin.

Most exciting of all, an expedition to Namibia earlier this year found living mantophasmatodid species in the tall grass atop the country's Brandberg Mountains. All specimens found so far are about two centimetres long. Analysis of stomach contents shows that they eat other insects. Three or four more species await description.

There are also fossil specimens. The species *Raptophasma*, described last year from an insect trapped in amber, shows that mantophasmatodids were present in Europe during the Eocene, about 45 million years ago. The living African species might be the remnants of a once-widespread group now perilously close to extinction, or they might still



be widespread in Africa.

Klass *et al.* suggest that the new order is most closely related to the stick insects, and to a group called the Grylloblattodea, or ice-crawlers. This group, known from about 25 species found on

mountaintops in North America and Asia, was the last new order, discovered in 1914. DNA sequencing now under way might help to pin down the Mantophasmatodea's exact position in the tree of life. **John Whitfield**

or neurodegeneration, as in Alzheimer's and Parkinson's diseases.

Song *et al.*'s discoveries² may lead to a better understanding of how neurogenesis could be stimulated. The authors isolated neural stem cells from the hippocampus of adult rats, and engineered the cells to express green fluorescent protein (GFP), allowing them and their progeny to be easily traced. After culturing the stem cells with astrocytes from the hippocampus of newborn rats, Song *et al.* found that the rate of neurogenesis increased more than eightfold. By contrast, GFP-labelled stem cells cultured with fibroblast cells or purified neurons did not change their rate of neurogenesis. Next, the authors showed that astrocytes derived from the adult hippocampus could also increase neurogenesis, albeit with less efficiency than hippocampal astrocytes from newborn rats. The astrocytes appeared to work by increasing the rate of proliferation of the stem cells and steering their progeny towards becoming neurons, rather than simply enhancing the survival of new neurons.

Is this feature common to all astrocytes? To find out, Song *et al.* purified astrocytes from the spinal cord of newborn and adult rats — a region that does not normally show neurogenesis *in vivo*. The spinal-cord astrocytes from newborns had only small effects on neurogenesis from hippocampal stem cells; adult spinal-cord astrocytes had no effect. So there is significant regional specificity in the ability of astrocytes to induce neurogenesis.

Song *et al.*'s demonstration that astrocytes can control the proliferation of hippocampal stem cells accords with previous studies that used stem cells from other brain regions. For example, astrocyte fragments can modulate the proliferation of stem cells from the cortex of developing rats³, and astrocyte monolayers can increase neurogenesis from adult subventricular-zone stem cells⁴. What is new and fascinating about Song *et al.*'s work is that it shows that astrocytes can direct the fate, as well as the proliferation, of hippocampal stem cells, inducing their progeny to become neurons rather than glia — and that this depends on using astrocytes from neurogenic brain regions.

One rather simple explanation for Song *et al.*'s regional-specificity results is that the spinal cord contains older astrocytes than the hippocampus, which constantly renews at least some of its astrocytes. (In fact, the neuron-producing hippocampal stem cells may even be astrocytes⁵.) In support of this idea, newly generated astrocytes can increase the growth of neuronal extensions (neurites), whereas older astrocytes cannot⁶. But whatever the explanation, it should be possible to dissect the mechanisms underlying the ability of astrocytes to control neurogenesis.

For example, researchers could use microarray technology to identify which genes are expressed differently in those astro-

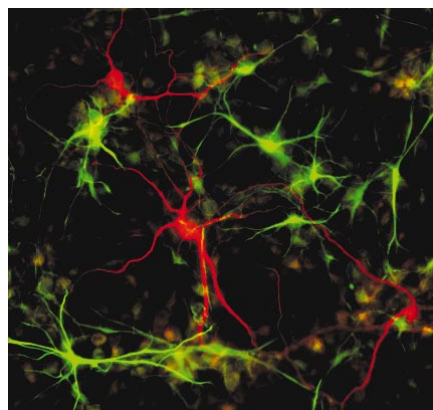


Figure 1 More than just glue. Glial cells such as astrocytes (green) were once thought to have a purely supportive role in the brain. But more recent work, including the paper by Song *et al.*², tells a different story. The image shows neurons (red) and astrocytes, derived from mouse neural stem cells.

cytes that can and cannot promote neurogenesis. Hot candidates might be genes encoding molecules similar to Noggin, which is known to be released from the ependymal cells lining the ventricles of the brain. Noggin can induce neurogenesis from subventricular-zone stem cells, both *in vitro* and following grafting into the brain, by inhibiting glial signalling pathways involving bone morphogenetic proteins (BMPs)⁷. As Song *et al.* mention, if neural stem cells are taken from the spinal cord and grown in culture, they can produce small numbers of neurons. This suggests that neurogenesis-inhibiting factors such as BMPs may be present in intact spinal cords. They may also be present in astrocyte cultures generated from the spinal cord, perhaps explaining why such astrocytes cannot promote the proliferation of hippocampal stem cells.

It is also interesting that stem cells generated from different regions of human and rodent embryonic brains retain their regional specificity. Those from the forebrain generate more neurons than those from the hindbrain and have distinct gene-expression patterns, even after long periods in culture^{8,9}. It remains to be seen whether this regional specificity of stem cells is due to their control by regionally specific astrocytes, but Song *et al.*'s paper makes it an intriguing possibility.

Another question is whether it might be possible to enhance brain repair simply by exposing stem cells that do not normally make neurons to 'neurogenic' astrocytes. An obvious way of finding out would be to transplant hippocampal astrocytes into a damaged spinal cord, and to see whether spinal-cord stem cells then produce neurons. Whatever the results, there will certainly be a complex interaction between the intrinsic gene-expression patterns of the stem cell, and environmental cues such as those produced by astrocytes. The 'nature versus nurture' debate has never been more interesting



100 YEARS AGO

In the *Proceedings* of the American Philosophical Society for December 1901 (vol. xl. No. 167), Mr. Percival Lowell refers at some length to the observations that led to the announcement in the Press that Mars had been signaling to the earth on a night in December 1900. It may be mentioned that the *original* dispatch read as follows:—"Projection observed last night over Icarium Mare, lasting seventy minutes." (Signed) "Douglas." In the present paper Mr. Lowell describes in detail some of the individual observations, and points out how the Flagstaff observations of 1894 showed that on general principles the Martian projections were most probably not due to the existence of mountain peaks. A close study of the surface markings led both Messrs. Lowell and Douglas to the result that these several projections were not caused by such permanent surface markings as mountains, but were the effect of clouds floating in the planet's atmosphere... Mr. Lowell, in his concluding remarks, says that the surface marking, Icarium Mare, is undoubtedly a great tract of vegetation, and the observation of December is completely explained if it be assumed that a cloud was formed over this region and rose to a height of thirteen miles, and then, traveling east by north at about twenty-seven miles an hour, passed over the desert of Aeria and there was dissipated. From *Nature* 1 May 1902.

50 YEARS AGO

A new field of the application of television was opened up about a year ago when, at the request of the admiralty, Marconi's Wireless Telegraph Co., Ltd., hurriedly assembled a television camera chain in an attempt to find the lost submarine *Affray*. A great deal of development work has since been carried out in co-operation with Siebe, Gorman and Co., Ltd.; and recently (April 17) a demonstration of the newly designed equipment was given in an experimental tank at the works of this Company. Among the special features of the apparatus shown were the use of the extremely sensitive image-orthicon camera tube, the enclosure of this camera and its associated components in a chamber capable of withstanding the water pressure prevailing at great depths, and the provision of remote-control facilities whereby the camera may be focused and directed by the operator who remains on the ship above. From *Nature* 3 May 1952.

— and is as complicated as ever. One thing, however, is clear. The humble astrocyte deserves a lot more attention.

Clive N. Svendsen is in the Stem Cell Research Program in Neuroscience, The Waisman Center, University of Wisconsin-Madison, 1500 Highland Avenue, Madison, Wisconsin 53705-2280, USA.

e-mail: svendsen@waisman.wisc.edu

1. Ullian, E. M., Sapperstein, S. K., Christopherson, K. S. & Barres, B. A. *Science* **291**, 657–661 (2001).

2. Song, H., Stevens, C. F. & Gage, F. H. *Nature* **417**, 39–44 (2002).
3. Temple, S. & Davis, A. A. *Development* **120**, 999–1008 (1994).
4. Lim, D. A. & Alvarez-Buylla, A. *Proc. Natl Acad. Sci. USA* **96**, 7526–7531 (1999).
5. Seri, B., Garcia-Verdugo, J. M., McEwen, B. S. & Alvarez-Buylla, A. A. *J. Neurosci.* **21**, 7153–7160 (2001).
6. Smith, G. M., Rutishauser, U., Silver, J. & Miller, R. H. *Dev. Biol.* **138**, 377–390 (1990).
7. Lim, D. A. *et al.* *Neuron* **28**, 713–726 (2001).
8. Ostenfeld, T. *et al.* *Dev. Brain Res.* **134**, 43–55 (2002).
9. Hitoshi, S., Tropepe, V., Ekker, M. & van der Kooy, D. *Development* **129**, 233–244 (2002).

Plant population biology

How to be invasive

Wim H. Van der Putten

Few clear answers have emerged from studies of the factors determining abundance of plants in particular settings. A new idea invokes the differing susceptibility of plant roots to damage from pathogenic soil microorganisms.

Why, in any particular ecosystem, are some plant species rare and others common? And why are some species highly successful in invading new territories whereas others never make it outside their own specific habitat? In tackling these questions, ecologists are trying to identify the primary factors that determine plant abundance. On page 67 of this issue¹, Klironomos presents three elegant experiments which show that natural plant abundance is related to the different rates at which soil pathogens develop on the roots of various species.

In his first experiment, Klironomos grew plants that are rare in semi-natural grasslands, and plants — originating from Eurasia — that have successfully invaded grasslands in North America since they were introduced more than a hundred years ago. He cultured examples of each plant species in pots containing soil from a grassland near the University of Guelph in Canada. After ten weeks, he removed all of the plant matter and put in new individuals of the same species. Then, after another ten weeks, he collected the plants and split the soil into two halves. In one half, the same plant species was grown for a third run; in the other half, examples of one of the other grassland species were grown.

After the third growth period, Klironomos

weighed the plants and compared the results from 'home' soil with those from soil in which other plant species had been grown. The rare plants produced less mass when grown in their home soil than in soil used to grow other species, whereas most of the invasive plants performed best when grown in their own soil. This is a remarkable result. From his initial observations, it suggested to Klironomos that roots of rare plant species might be accumulating all sorts of soil microbes, including soil pathogens, whereas invasive plants predominantly accumulate mycorrhizal root symbionts — which, rather than hindering plant growth, actively promote it.

That possibility is supported by the results of a second experiment, where Klironomos used a special sieving technique to separate the mycorrhizal soil organisms, which are symbiotic fungi with relatively large spores, from smaller soil microbes that may include root pathogens. Adding mycorrhizal fungi to sterilized soil planted with each of the rare and invasive plant species resulted in an increase in the weight of most plant species. This was expected — mycorrhizal fungi not only promote growth but have a fairly wide range of hosts², so that even non-native plant species can form associations with them³.

The more exciting result from this experi-

ment is that Klironomos found that adding the remaining soil microbes had an adverse influence on the rare plant species, but no effect on any of the invasive plants. A supplementary experiment showed that dominant fungi present on the plant roots may contribute to this effect — rare plants rapidly accumulated soil pathogens, whereas invasive plants did not.

These results tell us a great deal about plant invasiveness. In their new territories, plants are liberated from their native soil community which results in two clear benefits. First, the invading plants escape from their soil pathogens but don't encounter new, species-specific ones. Second, mycorrhizal fungi can associate with a broad range of plant species², meaning that root symbionts are likely to be available to the invader. This makes any plant species, even those that are rare in their native area, a potential invader.

The many previous studies on invasive plants have produced all sorts of results, so few general conclusions about the factors promoting invasiveness have emerged⁴. Factors such as absence of above-ground natural enemies and genetic variation of the founding populations have been considered. But this is the first time that the effects of escape from natural soil pathogens have been looked at.

We do not know if the invasive plant species used by Klironomos are sensitive to soil pathogens in their area of origin. But we do know that some plants that are controlled by soil pathogens in their home territories, such as American cherry (*Prunus serotina*)⁵, are highly invasive in new territories. We need further examples that include the responses of plant species in both their home and new territories to see how escape from specific soil pathogens may contribute to plant invasiveness.

In a final experiment, Klironomos grew a total of 61 species in a similar experimental set-up as before. By comparing the response of the plants to soils with different pre-growth histories, he found that rare plant species do better in the soil in which other species have grown than in their home soil, whereas it is the reverse for the more abundant plant species. The implication is that the rate at which plants accumulate soil pathogens can influence their potential abundance in the vegetation. Consequently, when plants escape from their soil pathogens, the patterns of abundance may drastically change. That happens in, for instance, the successional vegetation gradients of sand dunes⁶, in which infestation of pioneer plant species by specific soil pathogens facilitates replacement of the pioneers by later species.

The response of grassland plants to soil organisms can also be analysed from other perspectives. How, for instance, does a plant's sensitivity to the accumulation of soil pathogens relate to the plant's life history —

Figure 1 Common as ... this picture of a herbaceous plant community in the state of Washington shows an abundant species (*Plectritis congesta*, pink) and two rare ones (*Camassia* sp., violet; *Zigadenus* sp., yellow). Klironomos¹ digs into the reasons why some plants are abundant and others are not.



for example, whether it is an annual or a perennial. How do life histories evolve when you consider plant exposure to a whole range of below- or above-ground attackers? And are some plants more successful because they are protected against root pathogens by mycorrhizal fungi⁸ or some other means? The results of Klironomos¹ also show that we need to include the spatial and temporal dynamics of pathogen and symbiont development into models of plant–soil feedback⁹, and to relate these factors to plant abundance.

Finally, the concept of rarity used in this study¹ is of a different magnitude to that applying to some other ecosystems: for instance, the densities of certain tree species in tropical rainforests may be far less than one individual per hectare¹⁰. How soil pathogens contribute to biodiversity in circumstances

such as these remains an open question. ■

Wim H. Van der Putten is in the Department of Multitrophic Interactions, Netherlands Institute of Ecology (NIOO-CTO), PO Box 40, 6666 ZG Heteren, The Netherlands.

e-mail: putten@cto.nioo.knaw.nl

1. Klironomos, J. N. *Nature* **417**, 67–70 (2002).
2. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, London, 1997).
3. Callaway, R. M. *et al. Ecol. Lett.* **4**, 429–433 (2001).
4. Williamson, M. H. *Biological Invasions* (Chapman & Hall, London, 1996).
5. Packer, A. & Clay, K. *Nature* **404**, 278–281 (2000).
6. Van der Putten, W. H., Van Dijk, C. & Peters, B. A. M. *Nature* **362**, 53–56 (1993).
7. Van der Putten, W. H. *et al. Trends Ecol. Evol.* **16**, 547–554 (2001).
8. Newsham, K. K., Fitter, A. H. & Watkinson, A. R. *Trends Ecol. Evol.* **10**, 407–411 (1995).
9. Bever, J. D., Westover, K. M. & Antonovics, J. J. *Ecol.* **85**, 561–573 (1997).
10. Pitman, N. C. A. *et al. Ecology* **80**, 2651–2661 (1999).

Condensed-matter physics

Exciton developments

Ilias E. Perakis

The possibilities offered by Bose–Einstein condensation for investigating the quantum world continue to stretch the ingenuity of physicists. Quasiparticles known as excitons have become promising subjects for research.

The first Bose–Einstein condensate using atomic gases was created in 1995. This achievement, initially with rubidium, was made possible by the design of appropriate magnetic traps to hold the atoms, and the development of sophisticated cooling techniques¹. Work on producing a new kind of condensate — this time, using ‘excitons’ — has so far produced controversial results. But on page 47 of this issue, Butov *et al.*² provide firm evidence that high-density ‘lakes’ of cold excitons can be created in a semiconductor — a crucial step on the way to making an exciton Bose–Einstein condensate.

Excitons are quasiparticles: in semiconductors, electrons can be excited from the valence band to the conduction band using optical fields; this conduction electron is attracted, through the Coulomb interaction, to the positively charged hole left behind in the valence band. As a result, the electron and hole form a bound state called an exciton. Like a rubidium atom, this neutral bound complex can behave as a boson — a particle with integer spin which obeys Bose–Einstein statistics. When the temperature of a boson gas drops below a certain value, a large number of bosons ‘condense’ into a single quantum state — this is a Bose–Einstein condensate (BEC). All of these bosons then behave in exactly the same way, and quantum-mechanical effects become visible at a macroscopic level. Such collective boson behaviour gives rise to phenomena such as frictionless flow, or superfluidity, and quantum interference.

Several semiconductor systems have been investigated³ for evidence of an exciton BEC, with mixed results. Excitons have much lower mass than the atoms typically used to make BECs. This means that an exciton BEC can form at higher temperatures (although still only around the 1 K mark). The problem is that excitons exist only for a short time, just a few nanoseconds, before the electron and hole recombine. So it is difficult to create a ‘gas’ — a Bose gas — of excitons that is cold and dense enough to condense within this short time.

Butov’s group addressed this problem in earlier work⁴. They studied excitons con-

fined in a quantum well, an artificially grown two-dimensional structure of the semiconductors GaAs and AlGaAs. Such excitons can move freely only parallel to the quantum-well plane, and cannot move up and down in the perpendicular direction (the z-axis in Fig. 1a). Butov and colleagues designed a coupled quantum-well system in which ‘indirect’ excitons form when an electric field is applied in this perpendicular direction (Fig. 1b). This field pulls the oppositely charged electron and hole apart, trapping them tightly in the corners of their quantum wells. This reduces the spatial overlap of the electron and hole wavefunctions, with the result that the recombination rate is suppressed and the exciton lifetime increased. So indirect excitons have more time to cool down.

There is another advantage to using indirect excitons. Usually excitons are cooled through scattering with acoustic phonons (the quantized lattice vibrations in the semiconductor system). For quantum-well excitons, the cooling effect can happen up to a thousand times faster than for unconfined excitons. Butov *et al.*⁴ showed that the combination of these two effects can produce a gas of excitons that displays Bose statistics — the first step on the way to making an exciton BEC.

The next step is to confine the excitons sufficiently to make them condense. As the thermal energy decreases, a Bose gas condenses when the characteristic length at which quantum phenomena occur (the de Broglie wavelength) becomes comparable to the distance between particles. Confining a Bose gas in all directions decreases the interparticle distance and thus raises the value of the critical temperature at which condensation happens^{1,5}.

In their new work², Butov *et al.* observed the trapping of excitons in all three directions, within a radius of several micrometres (Fig. 1c). To achieve this, they took

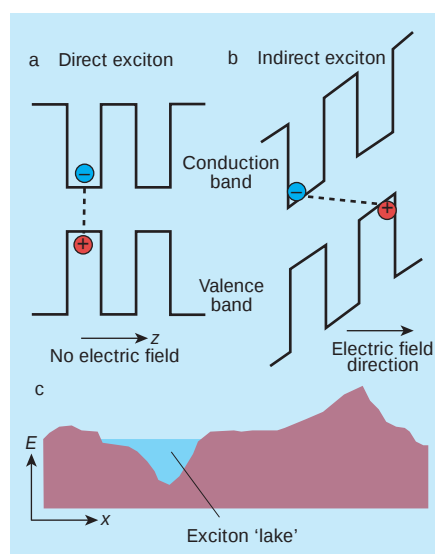


Figure 1 Towards a Bose–Einstein condensate (BEC) of excitons. a, An electron inside a quantum (potential) well in the conduction band of a semiconductor forms a ‘direct exciton’ with a corresponding positively charged hole in the valence band. b, In a coupled-well system, an ‘indirect exciton’ can form. Applying an electric field traps the electron and hole at the edges of their potential wells, making it less probable that the electron and hole recombine and destroy the exciton state. Because the lifetime of the indirect exciton is longer than that of a direct exciton, the chances of cooling and condensing indirect excitons into a BEC are better. c, Taking a step closer to an exciton BEC, Butov *et al.*² show that cold, dense ‘lakes’ of excitons form in natural wells of minimum potential energy (E). These wells arise from disorder in the system, which results in random fluctuations in the potential energy.

advantage of the natural potential wells that form within the plane of a quantum-well system. These wells are caused by the disorder in the system, which leads to random fluctuations in the potential energy. The excitons seek the regions of minimum potential energy, and eventually accumulate inside the bottom of the deepest potential trap. This trapping creates dense 'lakes' of cold excitons and, in combination with the small exciton mass, increases the condensation temperature to around 1 K — much higher than the critical temperature in atomic systems. Although Butov *et al.* haven't yet made an exciton BEC, the development of this trapping technique paves the way to creating exciton condensates in semiconductors.

Such condensates would open up exciting possibilities for observing and manipulating quantum properties, including coherence. On the technical side, researchers are mastering the art of growing high-quality GaAs/AlGaAs nanostructures, and advances in nanotechnology mean that arrays of exciton traps with novel quantum mechanical properties could now be engineered.

Quantum-well excitons are renowned for their linear and nonlinear optical properties⁶, and the quantum coherence in an exciton condensate could reveal novel optical effects and nonlinear optical dynamics, as all condensed excitons emit light in tandem. Ultimately, there could be applications in ultrafast digital logical elements and quantum computing. But in the meantime, an exciton BEC is an ideal system for studies of non-equilibrium quantum mechanics at the frontier of condensed-matter physics. ■

Ilias E. Perakis is on leave from Vanderbilt University at the Department of Physics, University of Crete, PO Box 2208, Heraklion, Crete, Greece.

e-mail: ilias.e.perakis@vanderbilt.edu

1. Pethick, C. J. & Smith, H. *Bose-Einstein Condensation in Dilute Gases* (Cambridge Univ. Press, 2002).
2. Butov, L. V., Lai, C. W., Ivanov, A. L., Gossard, A. C. & Chemla, D. S. *Nature* **417**, 47–52 (2002).
3. Griffin, A., Snoke, D. W. & Stringari, S. *Bose-Einstein Condensation of Excitons and Biexcitons* (Cambridge Univ. Press, 1995).
4. Butov, L. V. *et al.* *Phys. Rev. Lett.* **86**, 5608–5611 (2001).
5. Trauernicht, D. P., Wolfe, J. P. & Mysyrowicz, A. *Phys. Rev. B* **34**, 2561–2575 (1986).
6. Chemla, D. S. & Shah, J. *Nature* **411**, 549–557 (2001).

Developmental biology

Modular feedback

Christof Niehrs and Hans Meinhardt

To understand cell signalling during development, we need to know how whole signalling networks — not just their individual components — are regulated. Two new studies highlight this point.

In simplistic terms, living organisms can be broken down into modules — distinct ensembles of interacting elements that are used in a combinatorial fashion. Modules occur at all levels of biological organization and can encompass physical clusters (such as ribosomes and other multimolecule cellular machines) as well as processes (such as cell-to-cell signalling pathways). One kind of module is the 'synexpression group', a genetic network with an important role in animal development. Member genes of a synexpression group not only function in a common pathway, but are also expressed in the same tight temporal and spatial pattern.

Writing recently in *Nature Cell Biology*, Tsang *et al.*¹ and Fürthauer *et al.*² described such a group, in zebrafish embryos, that involves the fibroblast growth factor (FGF) protein and features in particular a newly discovered transmembrane protein, Sef. This protein turns out to be a negative-feedback regulator of the FGF cell-to-cell signalling pathway, highlighting an emerging theme — the superimposition of positive and negative feedback — in the composition of synexpression groups involved in signalling. This has implications for how each group functions as an integrated whole.

Cell-to-cell signalling by members of the FGF-protein family is involved in essentially every process in animal development, from formation of the embryonic axes to bone development, and from fruitflies to humans. So, to understand cell growth and specialization, it is clearly important to identify components of the FGF pathway. Tsang *et al.*¹ and Fürthauer *et al.*² have done just that with their discovery of a previously unknown zebrafish gene, *sef* (for 'similar expression to *fgf* genes').

As the name implies, the spatial pattern of expression of the *sef* gene in early zebrafish (and mouse) embryos is characteristic of that of two *fgf* genes, *fgf3* and *fgf8* (Fig. 1a, c). This is because the expression of *sef* is tightly regulated by FGF signalling. Moreover, the Sef protein also functions as a negative-feedback inhibitor in the FGF pathway. Its misexpression inhibits FGF signalling, and blocking its expression leads to too much signalling, causing characteristic malformations in zebrafish embryos. Other known negative-feedback inhibitors of the FGF pathway are the intracellular proteins encoded by the *sprouty2* and *sprouty4* genes. These genes are expressed in the same pattern as *sef*^{2,3} (Fig. 1b), as are *erm* and *pea3*, which encode gene-transcription factors and may mediate the nuclear response to FGF signalling⁴. Together, these genes form an FGF synexpression group (Fig. 1d).

Rather like bacterial 'operons', synexpression groups in animals are genetic modules in which the function and expression of different genes are closely correlated. Synexpression groups function in many different processes, from cell signalling and the cell-division cycle to protein secretion. Indeed, the coordinated expression of genes that act in the same cellular process is a widespread phenomenon in animals, as revealed by the expression clusters observed in large-scale microarray experiments, which detect the expression of many genes simultaneously.

Two other signalling-related synexpression groups that have been identified are those involving bone morphogenetic protein-4 (BMP4) and the Delta protein. In these groups, the genes in each signalling pathway are expressed in the same pattern as BMP4 or Delta because they are themselves regulated by that pathway⁵. As well as being used in a variety of developmental processes, each of these synexpression groups has been evolutionarily conserved from fish to frogs to mice. This begs the question of what exactly is so advantageous about these synexpression groups that maintains them under selective pressure. The answer may lie

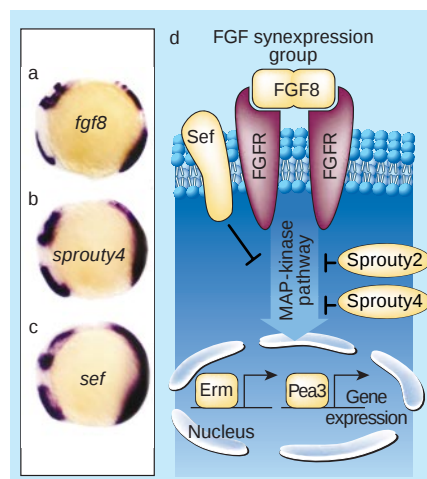


Figure 1 The fibroblast growth factor (FGF) synexpression group. a–c, The almost-identical patterns of expression (dark areas) of the *fgf8*, *sprouty4* and *sef*^{1,2} genes in zebrafish embryos. d, Simplified FGF8-triggered signalling pathway. Members of the FGF synexpression group (comprising proteins that are expressed in the same pattern as FGF and are involved in the same pathway) are shown in yellow. FGFR, FGF receptor. MAP kinases are enzymes that transmit the signal from the FGF receptor to the nucleus, where the transcription factors *Erm* and *Pea3* induce the expression of genes needed to carry out the pathway's functions. a–c are reproduced from ref. 2.

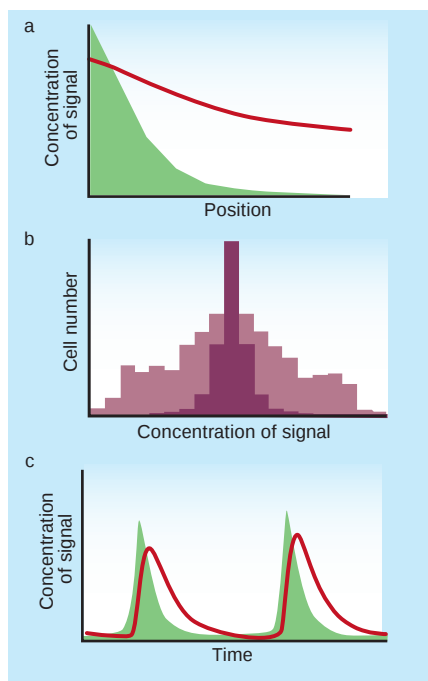


Figure 2 Negative feedback as an essential tool for generating appropriate behaviour of signalling systems. In a and c, the green regions represent the concentration of a signal (and hence of a positive-feedback regulator that is switched on by that signal). The red lines represent the concentration of a negative-feedback regulator that is also switched on by the signal. a, Negative feedback can regulate spatial patterns (such as gradients) of signal activation⁹. b, Negative feedback can reduce biological noise. The histogram shows how a cell population responds to a stimulus without (pale purple) or with (dark purple) negative feedback; negative feedback reduces the concentration range over which a signal fluctuates¹⁰. c, Negative feedback can lead to oscillations in signal concentration^{9,11}.

in how each group is regulated as a whole — that is, regulation at the ‘system’ level — particularly by negative feedback.

Signalling pathways typically contain both positive- and negative-feedback loops. Interestingly, a comparison of members of the FGF, BMP4 and Delta synexpression groups reveals that, although some encode proteins that positively reinforce signalling, most are negative-feedback inhibitors, as highlighted by *sef* and *sprouty* in the FGF pathway. Other examples from these synexpression groups are BAMBI, a pseudo-receptor for BMP that represses the BMP pathway⁶; Nrarp, a newly discovered intracellular protein that negatively regulates Notch, a receptor for Delta⁷; and the ESRs, a family of transcription factors that exert negative feedback on the expression of Delta⁸. Negative feedback can limit the absolute level of a signal and, together with positive self-regulation, can lead to spatial patterns of pathway activation (Fig. 2a)⁹. It can reduce the range over which the concentration of a signal fluctuates as a result of biological noise (Fig. 2b)¹⁰. And delayed negative feedback can lead to signal oscillations (Fig. 2c)¹¹.

It is not yet clear which of these effects is produced by *Sef*^{1,2}, or indeed exactly what it does at the molecular level. If, as Tsang *et al.*¹ discuss, FGF activates both a positive- and a negative-feedback loop, why don't these opposing influences just cancel each other out? Perhaps *Sef* is inhibitory only at high FGF concentrations, as would be expected if its role were to limit the maximum signal strength. If so, it will be important to find out how its concentration dependence is achieved. Alternatively, *Sef* may work to limit positive feedback, which might otherwise lead to a wave-like spread of pathway activation from cell to cell. However, theory predicts that such a negative

regulator would act over a longer range than FGF itself⁹, and this seems unlikely for a trans-membrane protein such as *Sef*. So *Sef* might be part of a receptor system that receives a signal that is different to, or modified from, that produced by FGF itself.

Future investigations must test whether synexpression groups involved in cell signalling are indeed modular circuits that are optimized to control properties at the system level. For example, it will be interesting to see whether *Sef* regulates the spread of FGF-induced signalling, the dynamics of the FGF network or its response to noise. To answer such questions, we will ideally require real-time read-outs of a signalling pathway while perturbing its components, backed up by network modelling and simulation. The quest to understand behaviour at the system level may lead to a new discipline at the interface of developmental biology, engineering and systems theory. ■

Christof Niehrs is in the Division of Molecular Embryology, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.

e-mail: niehrs@dkfz.de

Hans Meinhardt is at the Max-Planck-Institut für Entwicklungsbiologie, Spemannstrasse 37–39/IV, D-72076 Tübingen, Germany.

1. Tsang, M., Friesel, R., Kudoh, T. & Dawid, I. B. *Nature Cell Biol.* **4**, 165–169 (2002).
2. Fürthauer, M., Lin, W., Ang, S. L., Thisse, B. & Thisse, C. *Nature Cell Biol.* **4**, 170–174 (2002).
3. Fürthauer, M., Reifers, F., Brand, M., Thisse, B. & Thisse, C. *Development* **128**, 2175–2186 (2001).
4. Raible, F. & Brand, M. *Mech. Dev.* **107**, 105–117 (2001).
5. Niehrs, C. & Pollet, N. *Nature* **402**, 483–487 (1999).
6. Onichtchouk, D. *et al.* *Nature* **401**, 480–485 (1999).
7. Lamar, E. *et al.* *Genes Dev.* **15**, 1885–1899 (2001).
8. Jen, W.-C., Gawantka, V., Pollet, N., Niehrs, C. & Kintner, C. *Genes Dev.* **13**, 1486–1499 (1999).
9. www.eb.tuebingen.mpg.de/abt.4/meinhardt/theory.html
10. Becskei, A. & Serrano, L. *Nature* **405**, 590–593 (2000).
11. Elowitz, M. B. & Leibler, S. *Nature* **403**, 335–338 (2000).

Daedalus

Irregular solution

All chemists learn, as a sort of subtext to the topic, that a solution is a perfect mixture. Sugar dissolved in water, for example, has some sucrose molecules that are near each other, and some far apart, but the average does not vary. Demonstrators may homogenize a solution by shaking or stirring it. Daedalus's new imperfect solution will be much more regular.

He recalls nitrogen trichloride, that remarkably unstable molecule. Even a slight touch or slight heating will set it off: $2\text{NCl}_3 = \text{N}_2 + 3\text{Cl}_2$. This extreme sensitivity must depend on the ease with which two molecules can destroy each other to produce the highly stable dinitrogen molecule, N_2 . A single molecule could not form it, and should be more stable. So, says Daedalus, a hot, dilute solution of nitrogen trichloride should show a novel form of fluorescence. Whenever two molecules come together, they should react, liberating stable N_2 ; but on their own they should just stay in solution.

Now if two molecules happened to meet and decompose all at once, they would liberate a single photon with a wavelength of 260 nanometres in the far ultraviolet, a tricky spectral region. Probably, however, some molecular decomposition mechanism would liberate several less energetic photons in rapid succession. With luck, one of this shower would be visible, or at least perceptible to a near-UV detector.

When first made, and warmed to instability, Daedalus's non-statistical solution will fluoresce furiously as all the NCl_3 molecules near each other mutually annihilate, giving flashes of light. The fluorescence will then slow down to a dim ‘steady state’. Occasionally, molecules some distance apart may, by thermal wandering, make the erratic journey to a lethal encounter with each other. But most active molecules will be safely apart. A viscous solvent may be needed, to slow the rate at which single molecules make this lethal journey.

Fluorescence measurements will thus show whether single molecules ‘feel’ macroscopic viscosity; and if not, what viscosity they do feel. They should reveal the true laws of solution. The wandering of molecules will be open to study. Daedalus even hopes that his non-statistical solution will crystallize, or at least show a higher freezing-point than an ordinary one. Of course, it must be kept quite still. Any attempt to shake or stir it will bring distant molecules together. They will annihilate each other in a great glow.

David Jones

Rat navigation guided by remote control

Free animals can be 'virtually' trained by microstimulating key areas of their brains.

Procedures used to train laboratory animals often incorporate operant learning¹ paradigms in which the animals are taught to produce particular responses to external cues (such as aural tones) in order to obtain rewards (such as food). Here we show that by removing the physical constraints associated with the delivery of cues and rewards, learning paradigms based on brain microstimulation enable conditioning approaches to be used that help to transcend traditional boundaries in animal learning. We have used this paradigm to develop a behavioural model in which an experimenter can guide distant animals in a way similar to that used to control 'intelligent' robots.

Depending on the site of brain stimulation, an electrical stimulus can act as a cue or a reward²⁻⁴. Studies of these phenomena have generally been concerned with functional mechanisms of the nervous system⁵, and little thought has been given to the potential of behavioural paradigms constructed wholly around such focal brain stimulations. We used stimulation of the somatosensory cortical (SI) and medial forebrain bundle (MFB)³ as 'virtual' cues and rewards, respectively, delivered to freely roaming rats. We imposed behavioural contingencies so that an operator could accurately steer the animal, in real time, over any arbitrarily specified three-dimensional route and over a range of real-world terrains.

We implanted stimulating electrodes into the MFB of five rats; the same animals also received electrodes in the right and left SI whisker representations. We then mounted a backpack containing a microprocessor-based, remote-controlled microstimulator on each animal. This allowed the operator, using a laptop computer, to deliver brief trains of stimulus pulses (80 μ A; typically 10 biphasic pulses, each 0.5 ms, 100 Hz, per train) to any of the implanted brain sites from distances of up to 500 m away.

We trained the rats for navigation in 10 sessions, during which they learned to interpret remote brain stimulation as instructions for directing their trajectory of locomotion. In a figure-of-eight-shaped maze, the animals first learned to obtain periodic MFB rewards (0.3–3.0 Hz) by running forwards and turning correctly whenever left- or right-turning cues were issued. These cues were presented as a virtual 'touch' to the left or right whiskers by stimulating their respective cortical representations⁶. We then placed the animals in open environments that lacked the boundaries and fixed choice points of the maze. All rats generalized their responses to their new

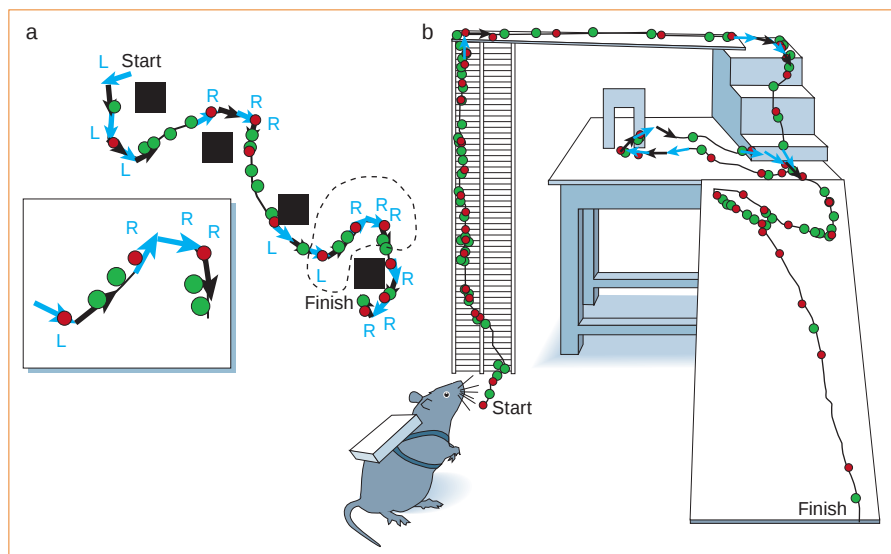


Figure 1 Examples of guided rat navigation using brain microstimulation. Sketches are constructed from digitized video recordings. Red dots indicate rat head positions at 1-s intervals; green dots indicate positions at which reward stimulations were administered to the medial forebrain bundle (MFB); blue arrows indicate positions at which right (R) and left (L) directional cues were issued; black arrows indicate positions 0.5 s after directional commands. **a**, Route followed by a rat guided through a slalom course. Inset, detail of the events that took place inside the dashed enclosure. **b**, Route taken by a rat guided over a three-dimensional obstacle course. The animal was instructed to climb a vertical ladder, cross a narrow ledge, descend a flight of steps, pass through a hoop and descend a steep (70°) ramp. Two rounds of high-density MFB stimulation were required to guide the rat successfully down the ramp, demonstrating the motivational qualities of MFB stimulation.

surroundings, running forwards and turning instantaneously on cue (Fig. 1a). They moved at speeds averaging 0.3 m s⁻¹ and worked continuously for periods of up to a 1-hour test limit.

Navigation over three-dimensional structures was achieved by incorporating a unique behavioural attribute of MFB stimulation that reflected the known 'priming' qualities of MFB stimulation^{7,8}. We found that MFB stimulation not only reinforced forward locomotion but also initiated and motivated further locomotion. Thus, an MFB reward itself served as an effective cue for rats to move forwards. Upon approaching objects such as a high step, 'forward' MFB stimulation would induce the rats to climb or descend from it. As a rule, the number of such stimulations required was proportional to the difficulty of negotiating the obstacle ahead (Fig. 1b). Superimposing 'forward' MFB stimulations onto the standard schedule was thus sufficient to steer the rats through a wide variety of complex, new and changing terrains.

Our rats were easily guided through pipes and across elevated runways and ledges, and could be instructed to climb, or jump from, any surface that offered sufficient purchase (such as trees). We were also able to guide rats in systematically exploring large, collapsed piles of concrete rubble,

and to direct them through environments that they would normally avoid, such as brightly lit, open arenas.

Our results show that 'virtual' learning, involving direct stimulation of the central substrates of cues and rewards, can effectively expand the scope of the operant method. Its chief benefit is its ability to dissociate explicit schedule variables such as cues and rewards from the physical variables that are normally associated with their delivery, freeing learning from the mechanical and parametric constraints that are imposed by particular physical settings. MFB reward stimulation is relatively non-satiating, and animals need not initiate consummatory behaviours to obtain such rewards. As virtual cues and rewards are perceived within a body-centred frame of reference, they may facilitate learning independently of the external environment. It may also be possible to increase the 'bandwidth' of conditionable information by stimulating multiple brain sites, thereby increasing the variety of reactions that can be elicited.

The specific behavioural model presented here — a guided animal — may have implications for new neurophysiological studies into directed animal navigation. MFB-based exploratory behaviours have already been shown to be useful in studying

hippocampal place cells⁹. Our model may also represent an extension of operant conditioning into useful real-world applications, such as search and rescue in areas of urban destruction and landmine detection. Combined with electronic sensing and navigation technology, a guided rat can be developed into an effective 'robot' that will possess several natural advantages over current mobile robots. Moreover, the ability to receive brain sensory activity remotely¹⁰ and interpret it accurately could allow a guided rat to function as both a mobile robot and a biological sensor.

Sanjiv K. Talwar*, **Shaohua Xu***,
Emerson S. Hawley*, **Shennan A. Weiss***,
Karen A. Moxon†, **John K. Chapin***

*Department of Physiology and Pharmacology,
State University of New York, Downstate Medical

Centre, 450 Clarkson Avenue, Brooklyn,
New York 11203, USA

e-mail: stalwar@netmail.hscbklyn.edu

†School of Biomedical Engineering, Drexel
University, 3141 Chestnut Street, Philadelphia,
Pennsylvania 19104, USA

1. Skinner, B. F. *The Behavior of Organisms: An Experimental Analysis* (Appleton-Century-Crofts, New York, 1938).
2. Loucks, R. B. *J. Comp. Psychol.* **16**, 439–444 (1933).
3. Olds, J. & Milner, P. J. *Comp. Physiol. Psychol.* **47**, 419–427 (1954).
4. Olds, M. E. & Fobes, J. L. *Annu. Rev. Psychol.* **32**, 523–574 (1981).
5. Doty, R. W. *Annu. Rev. Psychol.* **20**, 289–320 (1969).
6. Romo, R., Hernández, A., Zaino, A., Brody, C. & Lemus, L. *Neuron* **26**, 273–278 (2000).
7. Deutsch, J. A. *J. Theor. Biol.* **4**, 193–214 (1963).
8. Gallistel, C. R. *J. Comp. Physiol. Psychol.* **69**, 713–721 (1969).
9. Fukuda, M., Kobayashi, T., Bures, J. & Ono, T. *J. Neurosci. Methods* **44**, 121–131 (1992).
10. Hawley, E. S., Hargeaves, E. L., Kubic, J. L., Rivard, B. & Muller, R. U. *Hippocampus* (in the press).

Competing financial interests: declared none.

Development

Linguistic ability and early language exposure

For more than 100 years, the scientific and educational communities have thought that age is critical to the outcome of language learning^{1,2}, but whether the onset and type of language experienced during early life affects the ability to learn language is unknown. Here we show that deaf and hearing individuals exposed to language in infancy perform comparably well in learning a new language later in life, whereas deaf individuals with little language experience in early life perform poorly, regardless of whether the early language was signed or spoken and whether the later language was spoken or signed. These findings show that language-learning ability is determined by the onset of language experience during early brain development, independent of the specific form of the experience.

The ability to learn language, whether spoken or signed, declines with age^{3–6}. How the onset and type of the initial language experience contributes to this critical-period phenomenon is unclear. This question cannot be investigated by studying hearing individuals only, because the factors of age and experience are inseparable in these individuals — all hearing babies experience language from birth. But the question can be investigated by studying individuals who were born deaf, because they often do not experience any language until they are enrolled in special programmes^{7,8}. We therefore compared the language-learning capacities of deaf and hearing individuals as a function of early language experience.

We first investigated whether early experience of a spoken language could facilitate subsequent learning of a signed language. We tested two groups of adults who had

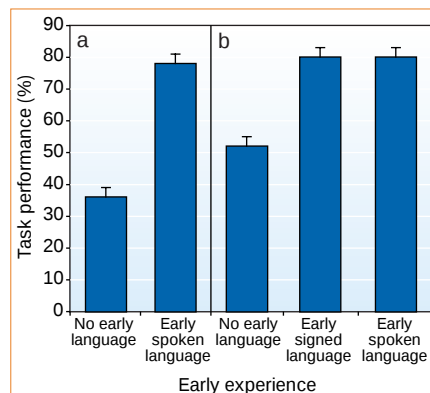


Figure 1 Effects of early experience on later language learning.

a, American Sign Language (ASL) performance of deaf adults who had experienced no language in early life, and of deaf adults who had experienced spoken language in early life. Subjects were tested using a task requiring recall of complex ASL sentences. **b**, English performance of deaf adults who had had no experience of language in early life, of deaf adults who had experienced ASL in infancy, and of hearing adults who had experienced a spoken language other than English in infancy. Subjects were tested using a task requiring judgements of whether complex English sentences given in print were grammatically correct; chance performance is 50%. Further details are available from the authors.

learned American Sign Language (ASL) at school between the matched ages of 9 and 15 years and who had used it for over two decades. One group ($n = 9$) was born hearing, had experienced spoken English in early life, and had later learned ASL after becoming profoundly deaf (≥ 90 decibels) as a result of viral infection; the second group ($n = 9$) was born profoundly deaf and had had little experience of language before being exposed to ASL in school (auditory speech-perception abilities were at chance levels even with hearing aids). Deaf adults who had little experience of language in early life showed low levels of ASL performance; in contrast, late-deafened adults showed high levels of ASL performance (Fig. 1; paired $t = 4.17$; d.f., 8; $P < 0.001$).

We next investigated whether early experience of a signed language facilitates subsequent learning of a spoken language. We tested three groups of adults who had learned English in school at comparable ages between 4 and 13 years and who had used it for over 12 years. One group ($n = 14$) was born profoundly deaf and had had little language experience before being exposed to ASL in school; the second group ($n = 13$) was born profoundly deaf and had experienced ASL in infancy; the third group ($n = 13$) was born hearing and had experienced various spoken languages in infancy (Urdu, French, German, Italian or Greek). Deaf and hearing adults who had experienced either a signed or a spoken language in early life showed similarly high levels of performance on the later learned language, English, whereas deaf adults who had little experience of language in early life showed low levels of performance (Fig. 1; $F_{2,37} = 11.32$, $P < 0.0001$).

Our results show that the ability to learn language arises from a synergy between early brain development and language experience, and is seriously compromised when language is not experienced during early life. This is consistent with current knowledge about how experience affects visual development in animals⁹ and humans¹⁰, and about learning and brain development in animals^{11,12}. The timing of the initial language experience during human development strongly influences the capacity to learn language throughout life, regardless of the sensorimotor form of the early experience.

Rachel I. Mayberry*, **Elizabeth Lock†**,
Hena Kazmi‡

*School of Communication Sciences and Disorders,
McGill University, 1266 Pine Avenue West,
Montreal, Quebec H3G 1A8, Canada
e-mail: rachel.mayberry@mcgill.ca

†Faculty of Medicine, University of Ottawa,
4418-501 Smyth Road, Ottawa,
Ontario K1H 8L6, Canada

‡School of Communication Sciences and Disorders,
University of Western Ontario, Elborn College,
London, Ontario N6G 1H1, Canada

1. Colombo, J. *Psychol. Bull.* **91**, 260–275 (1982).
2. Lenneberg, E. *Biological Foundations of Language* (Wiley, New York, 1967).
3. Johnson, J. & Newport, E. *Cogn. Psychol.* **21**, 60–69 (1989).
4. Newport, E. *Cogn. Sci.* **14**, 11–28 (1990).
5. Mayberry, R. I. & Eichen, E. J. *Mem. Lang.* **30**, 486–512 (1991).
6. Emmorey, K., Bellugi, U., Friederici, A. & Horn, P. *Appl. Psycholingu.* **16**, 1–23 (1995).
7. Mayberry, R. I. *J. Speech Hearing Res.* **36**, 51–68 (1993).
8. Mayberry, R. I. in *Child Neuropsychology* (eds Segalowitz, S. J. & Rapin, I.) (Elsevier, Amsterdam, in the press).
9. Wiesel, T. N. *Nature* **299**, 583–591 (1982).
10. Goldberg, M. C., Maurer, D., Lewis, T. L. & Brent, H. P. *Dev. Neuropsychol.* **19**, 55–81 (2001).
11. Greenough, W. T. & Black, J. E. in *Developmental Behavioral Neuroscience* (eds Gunna, M. R. & Nelson, C. A.) 155–200 (Erlbaum, Hillsdale, New Jersey, 1992).
12. Kolb, B., Forgie, M., Gibb, R., Gorny, G. & Rowntree, S. *Neurosci. Biobehav. Rev.* **22**, 143–159 (1998).

Competing financial interests: declared none.

Astroglia induce neurogenesis from adult neural stem cells

Hongjun Song^{*†}, Charles F. Stevens^{*} & Fred H. Gage[†]

^{*} Molecular Neurobiology Laboratory, Howard Hughes Medical Institute at the Salk Institute, and

[†] Laboratory of Genetics, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

During an investigation of the mechanisms through which the local environment controls the fate specification of adult neural stem cells, we discovered that adult astrocytes from hippocampus are capable of regulating neurogenesis by instructing the stem cells to adopt a neuronal fate. This role in fate specification was unexpected because, during development, neurons are generated before most of the astrocytes. Our findings, together with recent reports that astrocytes regulate synapse formation and synaptic transmission, reinforce the emerging view that astrocytes have an active regulatory role—rather than merely supportive roles traditionally assigned to them—in the mature central nervous system.

New cells are generated throughout the central nervous system (CNS) well into adulthood in all mammals, including humans^{1–4}. Adult neurogenesis occurs, however, only in two specific brain regions: the subventricular zone and the hippocampal subgranular zone^{1–4}. Outside these two regions, proliferating cells give rise to new glia but not to neurons in the intact adult CNS^{2,4–7}. Contrary to what one might anticipate, however, proliferating cells isolated from many regions in the adult brain^{8–14}, including non-neurogenic regions^{11–14}, can give rise to neurons both *in vitro*^{8–14} and after grafting back to neurogenic regions *in vivo*^{15,16}. These findings lead to an emerging view that neural stem cells may be widely distributed in the adult CNS and that the local environmental cues dictate their fate choice. Little is known at present about the mechanisms for the fate specification of adult neural stem cells³.

In this study, we exploited the advantages of cell culture systems to investigate the contributions made by different cell types to the fate specification of fibroblast growth factor-2 (FGF-2)-dependent stem cells derived from adult hippocampus. We demonstrate that mature astrocytes from postnatal hippocampus promote neurogenesis, whereas neurons increase oligogenesis by the same adult stem cells. Using a quantitative approach, we show that hippocampal astrocytes actively regulate adult neurogenesis both by instructing neuronal fate commitment and by promoting proliferation of adult neural stem cells. Moreover, the effects of astrocytes are regionally specified: astrocytes from adult hippocampus retain the potential to promote neurogenesis, but astrocytes from adult spinal cord do not. This finding of active roles for astroglia in adult neurogenesis is unexpected and raises the possibility that neuronal production in the mature brain is regulated, at least in part, by the regional properties of astrocytes.

Differentiation of adult stem cells

To mark the progeny of stem cells, we infected clonal derived neural stem cells isolated from the hippocampus of adult rat with a retrovirus engineered to express green fluorescent protein (GFP), and then selected GFP-expressing (GFP⁺) cells for this study^{10,17}. These FGF-2-dependent, GFP⁺ adult cells retain their stem cell properties. First, they can give rise to all three principal CNS cell types, as defined by cell-type-specific markers for neurons (Tuj1⁺, MAP2ab⁺), oligodendrocytes (GalC⁺, RIP⁺), or astrocytes (GFAP⁺), both *in vitro*^{10,17} (Fig. 1) and *in vivo* after grafting back to adult hippocampus¹⁵. Second, most of these cells actively proliferate in the presence of 20 ng ml^{−1} FGF-2, as shown by incorporation of bromodeoxyuridine (BrdU) during a 36-h period (97.9 ± 0.6%, *n* = 3; Fig. 1a). Third, these adult stem cells are

positive for nestin (99.3 ± 0.2%, *n* = 3; Fig. 1a)¹⁸, an immature cell marker, but negative for markers of many differentiated cell types (Fig. 1d).

These adult neural stem cells, unlike some neural stem cells derived from the developing brain¹⁹, appear to require exogenous factors to promote fate commitment to a neuronal lineage under our experimental conditions. This conclusion is based on the observation that the adult stem cells rarely differentiate spontaneously into neurons as defined by expression of MAP2ab, and only infrequently into oligodendrocytes (RIP⁺) or astrocytes (GFAP⁺), when cultured at a low density in a defined medium free of serum and FGF-2 (Fig. 1b, d). Previously, retinoic acid (0.5 μM) and fetal bovine serum (FBS, 0.5%) were shown to induce limited neuronal differentiation of these adult stem cells (Fig. 1d)¹⁷.

To dissect out the contributions made by different cell types to the fate choice of adult-derived stem cells, we cultured FGF-2-dependent GFP⁺ adult stem cells on a feeder layer of primary neurons and astrocytes derived from neonatal hippocampus in a defined serum-free medium. We found that within six days after co-culture the GFP⁺ adult stem cells gave rise to significant numbers of neurons (MAP2ab⁺), as well as to some oligodendrocytes (RIP⁺), astrocytes (GFAP⁺) and undifferentiated (nestin⁺) cells (Fig. 1c). This effect of environment on neurogenesis is robust: there was an approximate eightfold increase in the fraction of MAP2ab⁺ and GFP⁺ neurons in co-culture as compared with that for laminin- or poly-lysine-coated substrates in parallel cultures; however, fractions of GFAP⁺ and GFP⁺ astrocytes or RIP⁺ and GFP⁺ oligodendrocytes were not significantly changed (Fig. 1d). In contrast, a feeder layer of primary fibroblasts derived from rat skin²⁰ showed no significant effect on neurogenesis (Fig. 1d). Thus, factors from primary hippocampal neurons and/or astrocytes can specifically promote neurogenesis from these adult neural stem cells in the defined medium free of serum.

Differential effects of astrocytes and neurons

We next addressed the question of which cell type is necessary to promote neurogenesis—neurons, astrocytes or both? To determine whether neurons have a large role in promotion of neurogenesis, we prepared cultures enriched in primary neurons from neonatal hippocampus, in which over 70% of total cells were neurons (MAP2ab⁺) and less than 20% were astrocytes (GFAP⁺). When adult stem cells were plated in these neuron-enriched cultures, we did not observe a significant increase in neurogenesis as compared to stem cells plated on laminin-coated substrates (Fig. 2a, c), demonstrating that these primary neurons are not sufficient to

promote neurogenesis from adult stem cells under these conditions. Notably, the fraction of oligodendrocytes (RIP⁺ and GFP⁺) in neuron-enriched cultures increased significantly (Fig. 2a, c), indicating that primary neurons promote oligodendrocyte production from the adult neural stem cells²¹. When we used primary neurons specifically from neonatal dentate gyrus, instead of neurons from the whole hippocampus, we observed a similar increase in oligodendrocyte production without an increase in neurogenesis (data not shown).

We next examined whether astrocytes alone are sufficient to promote neurogenesis under our experimental conditions. We prepared astrocytes derived from neonatal hippocampus (GFAP⁺: 95 ± 3%; *n* = 3) without detectable primary oligodendrocytes (RIP⁺) or neurons (MAP2ab⁺) within six days under our co-culture conditions (see Methods). We observed a greater than tenfold increase in the percentage of neurons (MAP2ab⁺ and GFP⁺) produced from adult stem cells on a feeder layer of these astrocytes, as compared to those cultured on laminin-coated substrates in parallel cultures (Fig. 2b, c). Thus, primary hippocampal astrocytes alone are sufficient to promote neurogenesis from these adult stem cells. Astrocytes produce a variety of soluble and membrane-associated factors that influence multiple functions of the CNS²². To examine whether the effects of astrocytes on neurogenesis depend on diffusible or membrane-bound factors, we plated adult stem cells on coated substrates conditioned by primary astrocytes without contact²³, or directly on lightly fixed astrocytes²⁴. We found significantly more neurons (MAP2ab⁺ and GFP⁺) in both conditions than we observed on coated substrates in parallel cultures (Fig. 2c). These results indicate that both diffusible and membrane-bound factors are responsible for promoting neurogenesis from these adult stem cells.

Survival and proliferation

The effects of hippocampal astrocytes on neurogenesis of adult stem cells could be due to some combination of enhanced neuronal survival, increased proliferation of progenitors, and/or instruction of fate commitment by the adult stem cells to a neuronal lineage. We can determine by direct measurements whether astrocytes promote cell survival (measure the rate of cell death) or promote proliferation of progenitors (measure the rate of cell generation). To determine whether either of these two mechanisms has a role in increased neurogenesis, we established the time course of neurogenesis (defined by the expression of Tuj1, an early neuronal marker). As shown in Fig. 3a, there seems to be a constant production of Tuj1⁺ and GFP⁺ neurons in co-culture with astrocytes, and the effects of astrocytes were observed as early as 1 day after plating. In addition, BrdU labelling experiments showed that some cells were still dividing during the four-day experiments (Fig. 3a).

First, we asked whether a decrease in neural precursor or neuronal death could explain the increased number of neurons (Tuj1⁺) in co-culture with astrocytes, cells that are known to promote neuronal survival in long-term cultures²⁵. To quantify cell death of the progeny of adult stem cells, we used Hoechst 33342 (2 µg ml⁻¹) to visualize fragmented nuclei characteristics of apoptotic cells in live (Fig. 3b), rather than fixed, cultures; living cultures were used to avoid underestimating cell death owing to detachment of dying and/or dead cells from the culture substrates. This assay revealed a rather small percentage of dying/dead cells and no significant difference between different conditions at any time examined for these short-term cultures (Fig. 3b). Thus, better neuronal survival does not have a significant role in the increased net neurogenesis observed in these short-term co-cultures with astrocytes.

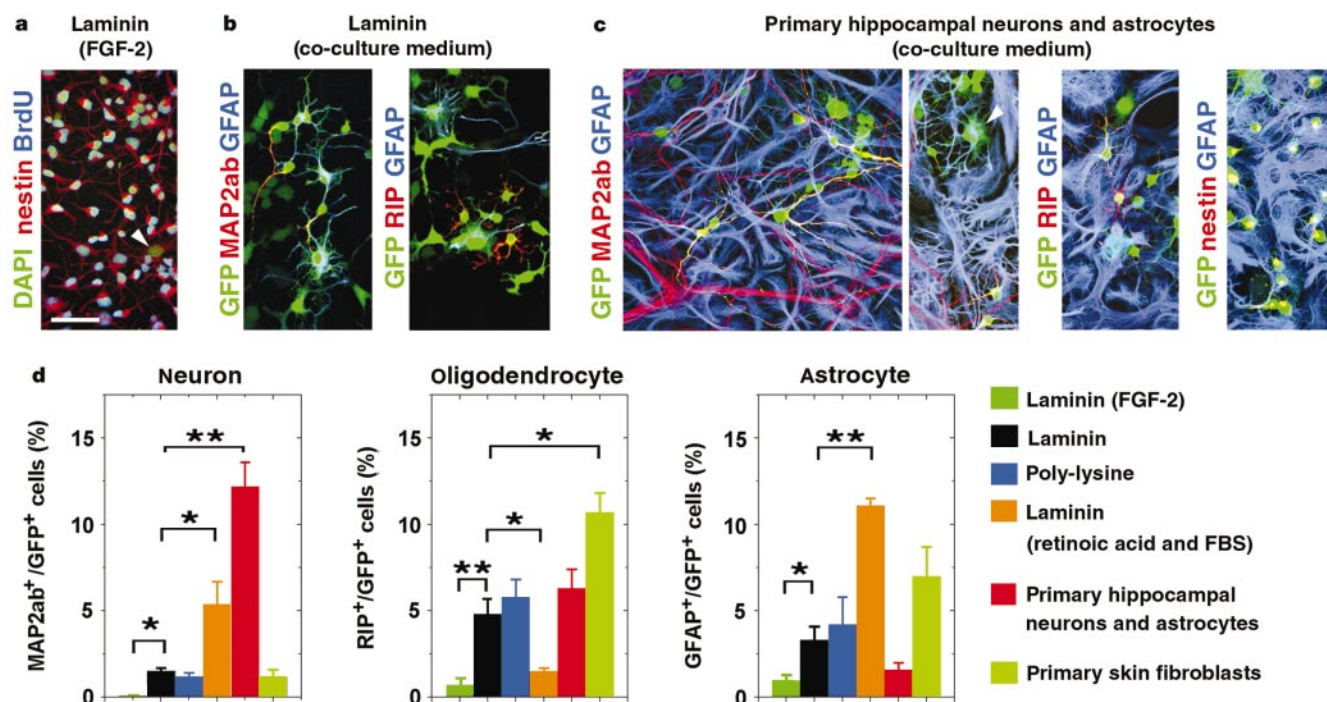


Figure 1 Differentiation of adult neural stem cells in a defined medium without serum. **a**, Proliferation of GFP⁺ adult stem cells in the presence of FGF-2. Cells were cultured on laminin-coated substrates in the presence of 2.5 µM BrdU and 20 ng ml⁻¹ FGF-2 for 36 h, and stained for BrdU and nestin. An arrowhead marks a nestin⁺ BrdU⁻ cell. Scale bar, 50 µm. DAPI, 4,6-diamidino-2-phenylindole. **b**, **c**, Differentiation of GFP⁺ adult stem cells cultured on laminin-coated substrates (**b**) or with neonatal hippocampal neurons and

astrocytes (**c**). Cells in six-day cultures were stained for markers for neurons (MAP2ab), astrocytes (GFAP), oligodendrocytes (RIP), or immature cells (nestin), respectively. An arrowhead marks a GFAP⁺ and GFP⁺ astrocyte. **d**, Quantification of six-day cultures. Data shown are mean values ± s.e.m. from four to eight experiments in parallel cultures. Significant differences from the control group (laminin) are marked with a single (*P* < 0.05, *t*-test) or double (*P* < 0.01, *t*-test) asterisk.

Next, we examined whether progenitor proliferation might contribute to the observed neurogenesis. Parallel cultures of GFP⁺ adult stem cells were incubated with BrdU (2.5 μ M) for 24 h on days 0, 1, 2 and 3 after plating, immediately followed by fixation and analysis of total cell numbers and BrdU incorporation. For GFP⁺ stem cells on laminin-coated substrates, we observed a gradual decline in the percentage of cells that were proliferating (GFP⁺ and BrdU⁺) during the four-day experiments (Fig. 3c). For those stem cells co-cultured with astrocytes, the fraction of GFP⁺ cells that divided during the first 24-h period was similar to that found on laminin-coated substrates, but became significantly higher at later time points (Fig. 3d). These observations indicate that factors from astrocytes do promote proliferation of adult neural progenitors. Consistent with results from BrdU labelling, we observed significantly more GFP⁺ cells in co-culture with astrocytes than on laminin-coated substrates after two days in culture (Fig. 3d). Time-lapse experiments confirmed that there was no significant difference in cell death between different culture conditions within the four-day experiments, and that there was an increase in proliferation for those co-cultured with astrocytes (Fig. 3e). Taken together, these results demonstrate that at least part of the effect of astrocytes on neurogenesis is due to proliferation of neural progenitors.

Neuronal fate commitment

Is the entire astrocytic effect on neurogenesis due to the proliferation of progenitors or do the astrocytes also instruct progenitors to

adopt a neuronal fate? To address this question, we developed a mathematical description of neurogenesis (as shown in Fig. 4a) that permits us to separate instructive effects from other possibilities. In our quantitative description of the process, we defined α_j as the proliferation rate (per day) of progenitors for the j th day in culture; β_j as the rate of conversion from progenitors (Tuj1⁻) to neurons (Tuj1⁺) during that day; and δ_j as the overall rate of cell death per day. The proliferation rate, α_j , is measured by the percentage of GFP⁺ cells that incorporate BrdU in a 24-h period (Fig. 3d). This measured rate was used to predict the total GFP⁺ cell numbers that had been counted in the same cultures. No significant difference was found for total GFP⁺ cell numbers between direct measurements (Fig. 4b, filled symbols) and theoretical prediction (Fig. 4b, open symbols). The death rate, δ_j , was measured directly by using

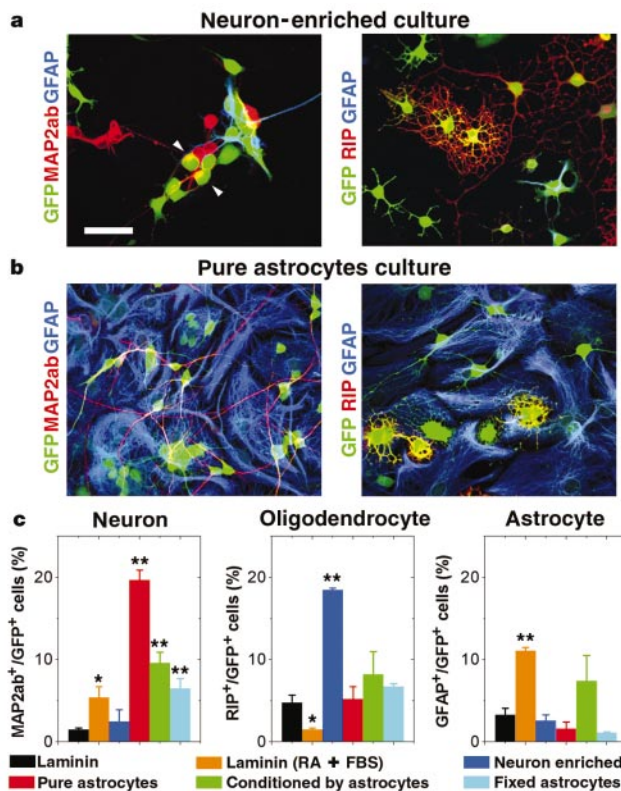


Figure 2 Distinct effects of primary astrocytes and neurons on the fate choice of adult neural stem cells. **a**, **b**, Differentiation of GFP⁺ adult neural stem cells in hippocampal neuron-enriched (**a**) or pure astrocyte cultures (**b**). Cells in six-day cultures were stained for MAP2ab, GFAP or RIP, respectively. Arrowheads mark GFP⁺ and MAP2ab⁺ neurons. Scale bar, 25 μ m. **c**, Quantification of six-day cultures. Data shown are mean values $\pm$ s.e.m. from four to eight independent experiments in parallel cultures. Significant differences from the control group (laminin) are marked with a single ($P < 0.05$, t -test) or double ($P < 0.01$, t -test) asterisk. RA, retinoic acid.

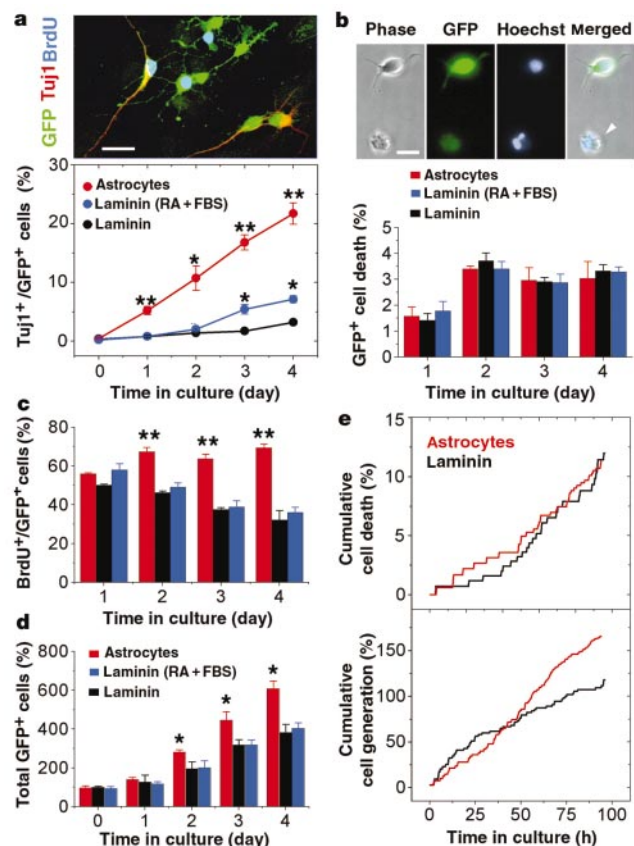


Figure 3 Effects of astrocytes on the survival, proliferation and neuronal fate commitment of adult neural stem cells. **a**, Time course of neurogenesis from adult neural stem cells. Cells cultured on hippocampal astrocyte, or laminin-coated substrates in the absence or presence of retinoic acid (RA, 0.5 μ M) and fetal bovine serum (FBS, 0.5%), were fixed at different time points and stained for an early neuronal marker, Tuj1. The top panel shows cells stained for Tuj1 and BrdU in the co-culture (four days) that was pulsed with BrdU (0.5 μ M) for 24 h at day 0. **b**, Cell death in different culture conditions as assessed daily by live staining with Hoechst 33342 (2 μ g ml⁻¹). The top panel shows example images in phase, GFP fluorescence, Hoechst 33342 fluorescence and merged. An arrowhead marks a dying cell. Scale bar, 2 μ m. **c**, Proliferation of GFP⁺ cells assessed by BrdU incorporation. Parallel cultures were incubated with BrdU (2.5 μ M) on different time points for 24 h, immediately followed by fixation and staining for BrdU. All data are mean values $\pm$ s.e.m. from four to six independent experiments in parallel cultures. Significant differences from the control group (laminin) are indicated by a single ($P < 0.05$, t -test) or double ($P < 0.01$, t -test) asterisk. **d**, The number of total GFP⁺ cells in different culture conditions. All data (same set as in **c**) were normalized to the total cell number on laminin-coated substrates at day 0 (4 h after plating). **e**, Cell generation and cell death of GFP⁺ cells under different conditions as assessed by time-lapse microscopy over four days. Images were acquired every 15 min. Death and generation of cells were quantified on each set of sequential images.

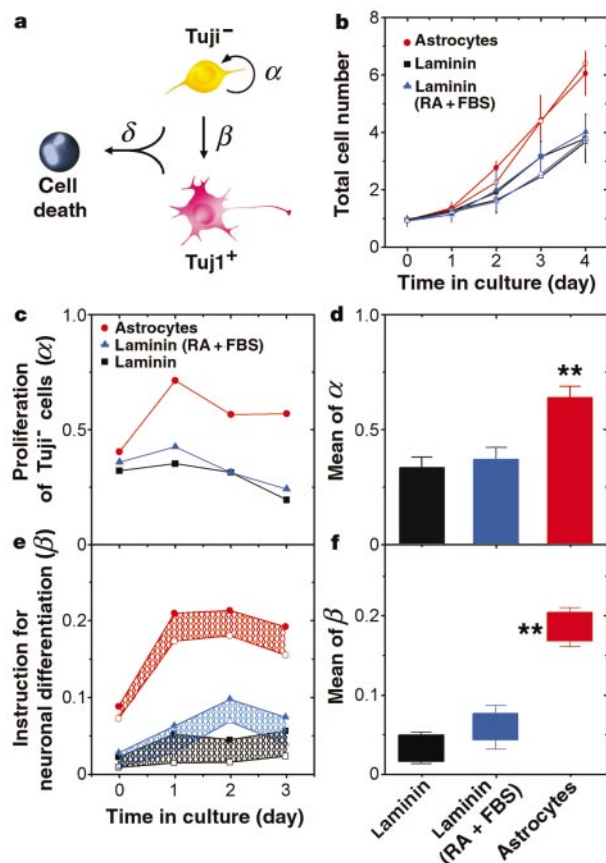


Figure 4 Astrocytes increase the rates of proliferation and neuronal fate commitment of adult neural stem cells. **a**, The states of progeny of adult neural stem cells in culture. α , cell division rate (per day) for non-neurons ($Tuji^{-}$); β , conversion rate of progenitors ($Tuji^{-}$) to neurons ($Tuji^{+}$); δ , cell death rate for all cells. **b**, Comparison of the total cell numbers from direct measurements (filled symbols) and theoretical prediction (open symbols) (see Methods). No significant differences were found between these two conditions in all culture conditions. **c, d**, The cell division rates for $Tuji^{-}$ cells (α). The time course of α (**c**) and the mean values $\pm$ s.e.m. for the last three days (**d**) is shown. **e, f**, The conversion rates of $Tuji^{-}$ cells to $Tuji^{+}$ cells (β). The upper limits (filled symbols) and lower limits (open symbols) for β in each condition are shown (**e**). The mean values $\pm$ s.e.m. for the last three days are indicated (**f**). Significant differences are indicated with a double asterisk ($P < 0.01$, t -test).

Hoechst 33342 dye in live cultures (Fig. 3b).

We could not, however, determine what fractions of the dying and dead cells were neurons. Counting the number of GFP^{+} neurons ($Tuji^{+}$) each day did permit us to estimate upper and lower limits on the conversion rate (β) of progenitors to neurons (Fig. 4e). These estimates were made with the limiting assumptions that either no neurons died or all cells that died were neurons. On the basis of these estimates, we conclude that adult stem cells proliferate about twice as rapidly on astrocytes (Fig. 4d) as those on laminin-coated substrate with or without the addition of retinoic acid ($0.5 \mu M$) and FBS (0.5%). Furthermore, and most importantly, the rate of conversion of progenitors to neurons is about six times higher for adult stem cells co-cultured with astrocytes (Fig. 4f). Thus, we are able to demonstrate quantitatively that hippocampal astrocytes increase the rates of both proliferation and neuronal fate commitment of FGF-2-dependent adult neural stem cells.

Developmental and regional specificity

These findings raised the question of whether distinct subtypes of glial cells are located in specific regions to regulate neurogenesis in the adult brain. To examine the effects of local environment on adult neurogenesis *in vivo*, we labelled proliferating cell populations, including the adult neural stem cells, in the dentate gyrus of adult hippocampus by pulsed injection of BrdU. Most BrdU-labelled cells were located near the inner molecular layer in intimate association with $GFAP^{+}$ astrocytes in the dentate gyrus of adult hippocampus, and were negative for any neuronal and glial markers at this stage (Fig. 5). Within a week of cell division, some progeny of proliferating cells began to express markers common for immature neurons and glia²⁰. Notably, as newly generated granule neurons mature, they migrate into the granule layer²⁶, where the cell bodies of granule neurons are densely packed with less exposure to $GFAP^{+}$ astrocytes. Recently, a vascular niche for adult hippocampal neurogenesis was identified²⁰. Astrocytes are well known to be in close association with vasculature^{27,28}, so they may provide an important part of the cellular elements that regulate adult neurogenesis in the dentate gyrus.

To examine whether adult astrocytes also regulate neurogenesis, we co-cultured GFP^{+} adult stem cells with astrocytes derived from hippocampus of adult rats (Fig. 6a). We found that significantly more GFP^{+} and $MAP2ab^{+}$ neurons are present in co-culture with adult hippocampal astrocytes as compared with those on coated substrates or skin fibroblasts (Fig. 6c); however, the promotion of neurogenesis by astrocytes derived from adult hippocampus is about half as effective as astrocytes derived from neonatal hippocampus (Fig. 6c). Thus adult hippocampal astrocytes retain the

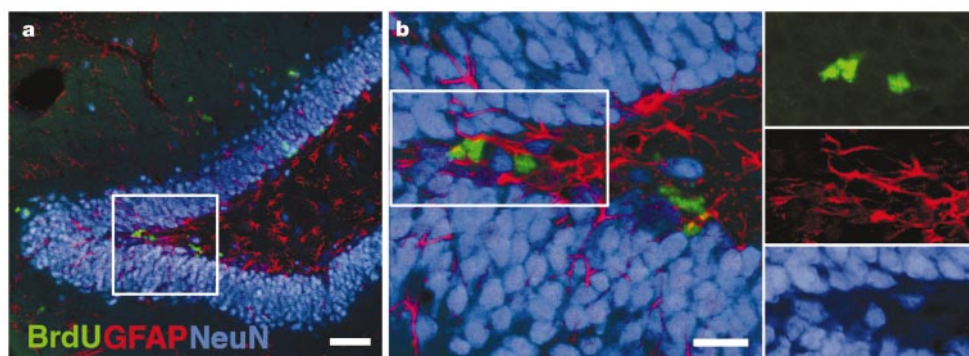


Figure 5 Intimate association of proliferating cells with $GFAP^{+}$ astrocytes in the dentate gyrus of adult hippocampus. Three-month-old rats were pulsed with BrdU (50 mg kg^{-1} , three times per day for three days). Tissues were fixed one day after the last BrdU injection

and stained for BrdU (green), astrocyte marker $GFAP$ (red) and neuronal marker $NeuN$ (blue). The region in the white box in **a** is shown at a higher magnification in **b**. Scale bar, $10 \mu m$ (**a**); $2 \mu m$ (**b**).

ability to promote neurogenesis from adult stem cells, but are less effective than the neonatal astrocytes, an effect consistent with the *in vivo* observation that the rate of neurogenesis decreases with age²⁹.

As noted above, active adult neurogenesis occurs in the subventricular zone and the subgranular zone of the dentate gyrus in hippocampus, but not in other regions of the intact CNS^{2,4-7}. Multipotent stem cells can, however, be isolated from many non-neurogenic regions¹¹⁻¹⁴, such as adult spinal cord^{11,12}. Our observation that hippocampal astrocytes can promote neurogenesis by adult stem cells raises the possibility that neurogenesis is absent in the adult spinal cord because the appropriate astrocytic signals are not present there. To examine this possibility, we co-cultured GFP⁺ adult stem cells with astrocytes from neonatal or adult spinal cord (Fig. 6b) and found that neonatal astrocytes from spinal cord produced some neuronal differentiation of these adult stem cells, but that adult astrocytes from spinal cord do support neurogenesis (Fig. 6c). In contrast, there is an increase in the fraction of GFAP⁺ and GFP⁺ astrocytes and RIP⁺ and GFP⁺ oligodendrocytes in co-culture with spinal cord astrocytes (data not shown). These results demonstrate that astrocytes from different CNS regions at several developmental stages exhibit different capabilities for regulating the fate choice of adult stem cells.

Discussion

In this study, we have examined the regulation of fate choices of FGF-2-dependent, adult-derived stem cells by different cell types that may exist in the niche of proliferative cells in adult brains. We demonstrated that primary neurons promote oligodendrocyte differentiation whereas astrocytes from postnatal hippocampus can actively regulate neurogenesis from adult neural stem cells.

Using a quantitative approach, we established that factors from astrocytes increase the rates of neuronal fate commitment and proliferation of progenitors by approximately sixfold and twofold, respectively (Fig. 4).

During development of the mammalian CNS, neurons and glial cells arise from multipotent progenitors in a stereotyped sequence through which neurons are generated first, primarily during the embryonic period, followed by glia, most of which differentiate after most neurons are generated^{3,27,28}. Astrocytes are known to support the proliferation, survival and maturation of developing neurons and neuroblasts that have already committed to neuronal lineages^{27,28,30}, as well as to stimulate neurogenesis from subventricular zone progenitors²⁴. In the mature mammalian brain, astrocytes constitute nearly half of the total cells, providing structural, metabolic and tropic support for neurons^{27,28}. Active, rather than supportive, roles for astrocytes in the adult CNS have been proposed only recently. For example, astrocytes have been reported to induce and/or stabilize CNS synapses^{31,32} and may be capable of integrating neuronal inputs and modulating synaptic activity³³⁻³⁶. In contrast to neurogenesis during development—when most astrocytes are not yet generated—adult neural stem cells reside in a significantly different environment. In the dentate gyrus of adult hippocampus, astrocytes are in intimate contact with proliferating cells *in vivo* (Fig. 5). Our results suggest active, rather than supportive, functions for astrocytes in adult brains: hippocampal astrocytes instruct the neuronal fate commitment of adult stem cells. Radial glia, which are a subpopulation of precursors generated before embryonic neurogenesis, were recently shown to give rise to neurons^{37,38}. It remains to be examined whether radial glia also directly regulate neurogenesis during development.

Active neurogenesis is limited to discrete regions of intact CNS of adult mammals¹⁻⁴. Local environments seem to dictate the fate choice of adult stem cells^{6,15,16,39}. For example, neural stem cells derived from adult hippocampus or spinal cord gave rise to neurons after grafting to the dentate gyrus, but not in the spinal cord¹⁶. In regions of adult mammalian neocortex that do not normally undergo any neurogenesis, endogenous neural precursors can differentiate *in situ* into neurons after induced synchronous apoptotic degeneration of corticothalamic neurons in the surroundings⁶. Consistent with *in vivo* observations⁵, we found that astrocytes from adult spinal cord, one of the non-neurogenic regions, are ineffective in promoting neurogenesis from adult stem cells. These results suggest that hippocampal astrocytes provide a unique niche for adult neurogenesis and present an intriguing possibility that the capability for adult neurogenesis might, in part, be due to certain signals provided by regionally specified astrocytes in the adult CNS. □

Methods

Cell cultures

The clonal derived stem cells isolated from hippocampus of adult Fisher 344 rats used in this study have been characterized previously^{10,17}. They were infected with retrovirus to express GFP, selected and propagated on laminin-coated flasks in DME/F12 medium containing N2 supplement, L-glutamine (2 mM) and FGF-2 (20 ng ml⁻¹) as previously described^{10,17}. For differentiation, the stem cells were trypsinized and plated in a serum-free defined medium (co-culture medium) containing MEM, N2 supplement, sodium pyruvate (1 mM), glucose (0.2 M) and L-glutamine (2 mM).

Primary astrocytes from hippocampus or spinal cord of postnatal day 0 (P0) or adult (older than six weeks) rats were prepared essentially as described^{23,40-42}. Importantly, confluent astrocytes were treated with cytosine arabinoside (20 μ M) for 72 h to eliminate proliferating cells, followed by recovery in fresh medium for 24 h. These steps may be repeated to further eliminate other cell types^{41,42}. Under these conditions, most of the cells were GFAP⁺ and/or S-100 β ⁺ astrocytes with little contamination by RIP⁺ or MAP2ab⁺ cells within 6 days under the co-culture conditions. We prepared primary neurons from P0 hippocampus as described previously⁴⁰. Primary neurons were plated at 10,000 cells cm⁻² on a confluent feeder layer of astrocytes from P0 hippocampus in the co-culture medium. For neuron-enriched cultures, P0 neurons from whole hippocampus, or only from dentate gyrus, were plated at 100,000 cells cm⁻² on poly-lysine-coated coverslips and cultured for two days in MEM containing N2 supplement, horse serum (10%), sodium pyruvate (1 mM), glucose (0.2 M) and cytosine arabinoside (5 μ M). We then changed the medium

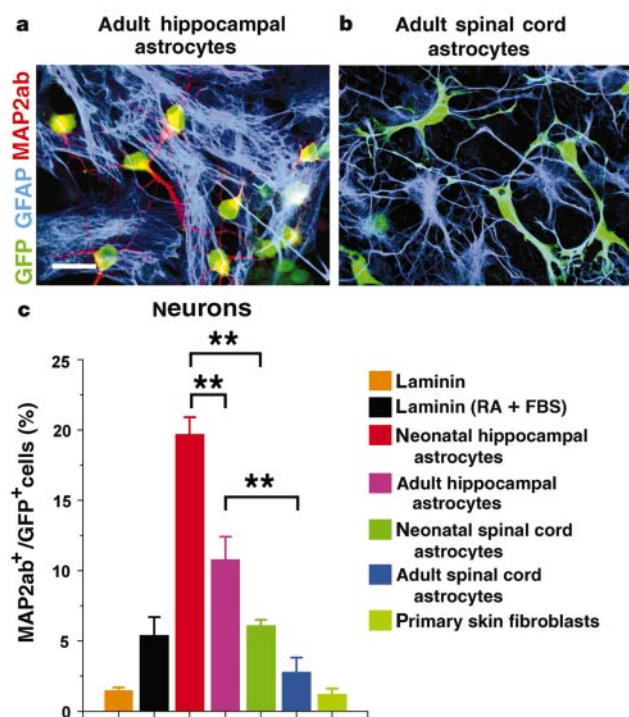


Figure 6 Mature astrocytes from adult hippocampus, but not adult spinal cord, promote neurogenesis from adult stem cells. **a, b**, Differentiation of GFP⁺ adult neural stem cells in co-culture with astrocytes derived from adult hippocampus (**a**) or adult spinal cord (**b**). Cells in six-day cultures were stained for MAP2ab and GFAP. Scale bar, 5 μ m. **c**, Quantification of the percentage of MAP2ab⁺ and GFP⁺ neurons in different conditions (six-day culture). Data shown are mean values $\pm$ s.e.m. from four to eight experiments in parallel cultures. Significant differences between results for astrocytes from hippocampus and spinal cord are indicated by a double asterisk ($P < 0.01$, *t*-test).

to fresh co-culture medium. These cultures survived for more than two weeks with over 70% of total cells as MAP2ab⁺ neurons and less than 20% as GFAP⁺ astrocytes. Primary skin fibroblasts from Fischer 344 rats²⁰ were cultured in DMEM containing 10% FBS and did not contain any detectable MAP2ab⁺, GFAP⁺ or RIP⁺ cells.

For co-culture experiments, substrate cells were cultured on poly-lysine- and collagen-coated coverslips until they were confluent. Fixed astrocyte feeder layers were prepared by light fixation with cold 70% ethanol at −20 °C for 30 min²⁴. For co-cultures using Banker culture, astrocytes were cultured on the bottom of 24-well plates, while laminin-coated coverslips were placed in the same well without contact²⁵. Substrate cells or coated coverslips were incubated in co-culture medium overnight before GFP⁺ stem cells were plated at 10,000 cells cm^{−2} in parallel for all culture conditions. Cultures were fed every other day. We added BrdU (2.5 μM) in some cultures to label dividing cells.

Immunocytochemistry and quantification

Cultures were processed for immunocytochemistry as described previously^{10,17}. We used the following primary antibodies: BrdU (1:400, rat; Accurate), nestin (1:1,000, mouse; Pharmingen), α-type III β-tubulin (Tuj1, 1:1,000, mouse; BAVCO), MAP2ab (1:250, mouse; Sigma), RIP (1:25, mouse; Hybrodoma Bank), S-100β (1:500, rabbit; Sigma) and GFAP (1:500, guinea-pig; Advance Immuno or 1:1,000, rabbit; Dako). The detection of BrdU in cultured cells required treatment in 1M HCN at 37 °C for 30 min¹⁰. Images were taken on a Bio-Rad confocal system at an original magnification of ×40 or ×63 with individual filter sets for each channel, and were assembled in PhotoShop.

Fluorescently labelled cells were visualized using confocal microscopy. Positive cells were quantified in at least 20 fields systematically across the coverslips from four to eight independent experiments of parallel cultures. Statistical analysis was performed with Prism.

Time-lapse recordings

GFP⁺ stem cells were plated onto different substrates and cultured for four days in recording chambers with 5% CO₂ at 37 °C on a microscope stage. Both phase and fluorescent images were captured every 15 min. Cell death and cell generation can be clearly identified on sequential images and thus analysed for a period of four days.

Quantitative analysis of neurogenesis from adult stem cells

The state diagram describing our experiments is shown in Fig. 4a. N_{j+1} is the total number of cells present at the start of the $(j+1)$ day in culture, and n_{j+1} is the total number of neurons (Tuj1⁺ cells) present at that time. α_j is the cell division rate for non-neurons (Tuj1[−] cells) on the j th day; β_j is the conversion rate of progenitors to neurons for that day; and δ_j is the death rate of all cells. For example, the number of cells that die during the j th day in culture is $\delta_j N_j$ —the other rates are defined in an analogous way. The total number of cells present at the start of the $(j+1)$ day, then, is given by the equation $N_{j+1} = (1 + \alpha_j) \times (N_j - n_j) + n_j - \delta_j N_j$. The first term on the right gives the contribution to the number due to progenitor division, the second term is the number of neurons present (assumed not to divide), and the third term specifies the contribution due to cell death. The cell division rate, α_j , is also given by the equation $N_{j+1}^* = 2\alpha_j(N_j - n_j)$, where N_{j+1}^* is the number of BrdU-labelled cells at the start of the $(j+1)$ day. Finally, the number of neurons present at the start of the $(j+1)$ day is given by the equation $n_{j+1} = n_j + \beta_j(N_j - n_j) - \delta_j^* n_j$, where δ_j^* is the death rate of neurons and the quantity $\delta_j^* n_j$ gives the number of neurons that die during the j th day. This equation states that the number of neurons present at the start of the $(j+1)$ day is made up of the ones present at the start of the j th day (n_j), plus the number of neurons produced by conversion of progenitors, $\beta_j(N_j - n_j)$, minus the number of neurons that died. We do not know this death number, but the inequality $\delta_j N_j \geq \delta_j^* n_j \geq 0$ places limits on the neuronal death. These equations, together with the measured numbers of total GFP⁺ cells, neurons, BrdU-labelled cell numbers, and cell death each day permit us to obtain α_j , β_j and δ_j (see Fig. 4).

In vivo BrdU labelling

Three-month-old Fischer 344 rats were injected intraperitoneally nine times with BrdU (50 mg kg^{−1}) over the course of three days, and tissues were fixed one day after the last injection. Tissues were processed for immunostaining as described previously²⁰.

Received 22 November 2001; accepted 20 March 2002.

1. Temple, S. & Alvarez-Buylla, A. Stem cells in the adult mammalian central nervous system. *Curr. Opin. Neurobiol.* **9**, 135–141 (1999).
2. Gage, F. H. Mammalian neural stem cells. *Science* **287**, 1433–1438 (2000).
3. Temple, S. The development of neural stem cells. *Nature* **414**, 112–117 (2001).
4. Rakic, P. Adult neurogenesis in mammals: an identity crisis. *J. Neurosci.* **22**, 614–618 (2002).
5. Horner, P. et al. Proliferation and differentiation of progenitor cells throughout the intact adult rat spinal cord. *J. Neurosci.* **20**, 2218–2228 (2000).
6. Magavi, S. S., Leavitt, B. R. & Macklis, J. D. Induction of neurogenesis in the neocortex of adult mice. *Nature* **405**, 892–893 (2000).
7. Kornack, D. R. & Rakic, P. Cell proliferation without neurogenesis in adult primate neocortex. *Science* **294**, 2127–2130 (2001).
8. Reynolds, B. A. & Weiss, S. Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science* **255**, 1707–1710 (1992).
9. Lois, C. & Alvarez-Buylla, A. Proliferating subventricular zone cells in the adult mammalian forebrain can differentiate into neurons and glia. *Proc. Natl Acad. Sci. USA* **90**, 2074–2077 (1993).
10. Palmer, T. D., Takahashi, J. & Gage, F. H. The adult rat hippocampus contains primordial neural stem cells. *Mol. Cell. Neurosci.* **8**, 389–404 (1997).

11. Shihabuddin, L. S., Ray, J. & Gage, F. H. FGF-2 is sufficient to isolate progenitors found in the adult mammalian spinal cord. *Exp. Neurol.* **148**, 577–586 (1997).
12. Kehl, L. J., Fairbanks, C. A., Laughlin, T. M. & Wilcox, G. L. Neurogenesis in postnatal rat spinal cord: a study in primary culture. *Science* **276**, 586–589 (1997).
13. Palmer, T. D., Markakis, E. A., Willhoite, A. R., Safar, F. & Gage, F. H. Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS. *J. Neurosci.* **19**, 8487–8497 (1999).
14. Kondo, T. & Raff, M. Oligodendrocyte precursor cells reprogrammed to become multipotential CNS stem cells. *Science* **289**, 1754–1757 (2000).
15. Suhonen, J. O., Peterson, D. A., Ray, J. & Gage, F. H. Differentiation of adult hippocampus-derived progenitors into olfactory neurons *in vivo*. *Nature* **383**, 624–627 (1996).
16. Shihabuddin, L. S., Horner, P. J., Ray, J. & Gage, F. H. Adult spinal cord stem cells generate neurons after transplantation in the adult dentate gyrus. *J. Neurosci.* **20**, 8727–8735 (2000).
17. Takahashi, J., Palmer, T. D. & Gage, F. H. Retinoic acid and neurotrophins collaborate to regulate neurogenesis in adult-derived neural stem cell cultures. *J. Neurobiol.* **38**, 65–81 (1999).
18. Lendahl, U., Zimmermann, L. B. & McKay, R. D. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**, 585–595 (1990).
19. Qian, X. et al. Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* **28**, 69–80 (2000).
20. Palmer, T. D., Willhoite, A. & Gage, F. H. Vascular niche for adult hippocampal neurogenesis. *J. Comp. Neurol.* **425**, 479–494 (2000).
21. Barres, B. A. & Raff, M. C. Axonal control of oligodendrocyte development. *J. Cell Biol.* **147**, 1123–1238 (1999).
22. Ridet, J. R., Malhotra, S. K., Privat, A. & Gage, F. H. Reactive astrocytes: cellular and molecular cues to biological function. *Trends Neurosci.* **20**, 570–577 (1997).
23. Goslin, K., Asmussen, H. & Banker, G. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 339–370 (The MIT Press, Cambridge, Massachusetts, 1998).
24. Lim, D. A. & Alvarez-Buylla, A. Interaction between astrocytes and adult subventricular zone precursors stimulates neurogenesis. *Proc. Natl Acad. Sci. USA* **96**, 7526–7536 (1999).
25. Banker, G. A. Trophic interactions between astroglial cells and hippocampal neurons in culture. *Science* **209**, 809–810 (1980).
26. Cowan, W. M., Stanfield, B. B. & Kishi, K. in *Current Topics in Developmental Biology* (ed. Hunt, K.) 103–157 (Academic, New York, 1980).
27. Nicholls, J. G., Martin, A. R. & Wallace, B. G. *From Neuron To Brain* 3rd edn, 149–152 (Sinauer, Sunderland, Massachusetts, 1992).
28. Kandel, E. R., Schwartz, J. H. & Jessell, T. M. *Principle of Neural Science* (McGraw-Hill, New York, 2000).
29. Kuhn, H. G., Dickinson-Anson, H. & Gage, F. H. Neurogenesis in the dentate gyrus of the adult rat: age-related decrease of neuronal progenitor proliferation. *J. Neurosci.* **16**, 2027–2033 (1996).
30. Pixley, S. K. CNS glial cells support *in vitro* survival, division, and differentiation of dissociated olfactory neuronal progenitor cells. *Neuron* **8**, 1191–1204 (1992).
31. Ullian, E. M., Sapperstein, S. K., Christopherson, K. S. & Barres, B. A. Control of synapse number by glia. *Science* **291**, 657–661 (2001).
32. Mauch, D. H. et al. CNS synaptogenesis promoted by glia-derived cholesterol. *Science* **294**, 1354–1357 (2001).
33. Haydon, P. G. GLIA: listening and talking to the synapse. *Nature Rev. Neurosci.* **2**, 185–193 (2001).
34. Lino, M. K. et al. Glia-synapse interaction through Ca²⁺-permeable AMPA receptors in Bergmann glia. *Science* **292**, 926–929 (2001).
35. Oliet, S. H. et al. Control of glutamate clearance and synaptic efficacy by glial coverage of neurons. *Science* **292**, 923–926 (2001).
36. Smit, A. B. et al. A glia-derived acetylcholine-binding protein that modulates synaptic transmission. *Nature* **411**, 261–268 (2001).
37. Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).
38. Miyata, T., Kawaguchi, A., Okano, H. & Ogawa, M. Asymmetric inheritance of radial glial fibers by cortical neurons. *Neuron* **31**, 727–741 (2001).
39. Lim, D. A. et al. Noggin antagonizes BMP signalling to create a niche for adult neurogenesis. *Neuron* **28**, 713–726 (2000).
40. Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).
41. Noble, M. & Mayer-Proschel, M. in *Culturing Nerve Cells* (eds Banker, G. & Goslin, K.) 499–544 (The MIT Press, Cambridge, Massachusetts, 1998).
42. Schwartz, J. P. & Wilson, D. J. Preparation and characterization of type 1 astrocytes cultured from adult rat cortex, cerebellum, and striatum. *Glia* **5**, 75–80 (1992).

Acknowledgements

We thank M. L. Gage and J. Sullivan for comments. This work was supported in part by grants from the National Institutes of Health, Howard Hughes Medical Institute, Christopher Reeve Paralysis Foundation, the National Institute of Aging, the Michael J. Fox Foundation, Project ALS and The Lookout Fund. H.S. is an associate and C.F.S. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.F.S. (e-mail: stevens@salk.edu) and F.H.G. (e-mail: gage@salk.edu).

The confinement of Neptune's ring arcs by the moon Galatea

Fathi Namouni & Carolyn Porco

Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80302, USA

Neptune has five narrow ring arcs, spanning about 40 degrees in longitude, which are apparently confined against the rapid azimuthal and radial spreading that normally results from inter-particle collisions. A gravitational resonance based on the vertical motion of the nearby neptunian moon Galatea was proposed^{1,2} to explain the trapping of the ring particles into a sequence of arcs. But recent observations^{3,4} have indicated that the arcs are away from the resonance, leaving their stability again unexplained. Here we report that a resonance based on Galatea's eccentricity is responsible for the angular confinement of the arcs. The mass of the arcs affects the precession of Galatea's eccentric orbit, which will enable a mass estimate from future observations of Galatea's eccentricity.

In the post-Voyager model², the moon Galatea produces a set of equilibrium points rotating at a constant rate (the pattern speed) that depends on the moon's mean motion and its vertical frequency about the planet's Laplace plane. However, as these corotation sites are potential maxima, dissipative inter-particle collisions tend to remove arc particles from the corotation sites, spreading the arcs azimuthally. Galatea can counterbalance the disruptive collisions by forcing the arc particles' orbits to be eccentric, thereby acting as an external source of energy. Before the recent observations, the arcs fell almost exactly at the location of the outer 86:84 corotation inclination resonance (CIR) of Galatea (Table 1, Fig. 1a,b) and within the 6 km width of its 43:42 Lindblad resonance (LR1) that forces the arc particle orbits to be eccentric. The 1998 observations of Neptune's arcs, however, yielded new mean motion measurements, $820.1122 \pm 0.0003 \text{ deg d}^{-1}$ (ref. 5) and $820.1135 \pm 0.0009 \text{ deg d}^{-1}$ (ref. 4) that differ from the previously inferred one $820.1185 \pm 0.0004 \text{ deg d}^{-1}$ (refs 2, 6) by about $-5 \times 10^{-3} \text{ deg d}^{-1}$, equivalent to an offset of 0.3 km. The new mean motion had in fact appeared previously in the data fits⁶ obtained by combining ground-based and Voyager observations but was rejected in favour of the old value because the former displaces the arcs outside the CIR islands (Fig. 1c, d) where angular confinement is lost. A closer analysis of this offset shows that it cannot be accounted for by fine-tuning the system's parameters. More precisely, displacing the CIR from its current location by modifying Neptune's radius and quadrupole moment either implies large changes of order 25% in these parameters or introduces erroneous systematic offsets to the motion of Neptune's satellites. Alternatively, the CIR location can also be displaced by separately decreasing the regression rate of Galatea's mode under the action of a massive arc system. However, the required change, $-0.185 \text{ deg d}^{-1}$, necessitates a mass at least ten

times larger than Galatea's. Finally, the CIR can be widened by increasing Galatea's mass; but in this case, the satellite must be at least three times more massive than the current estimate. This is inconsistent with the observed eccentric distortions forced on the arc system by Galatea².

The nearby corotation eccentricity resonance (CER) would be a candidate for the arcs' confinement were it not located 2 km inside the arcs' orbit (Table 1). The CER offers 43 potential maxima where arc particles can be trapped, with an angular extension of 8.37 deg , rotating at the pattern speed $n_{\text{CER}} = n_G - \kappa_G/43 = 820.1481 \text{ deg d}^{-1}$ where n_G and κ_G are, respectively, Galatea's mean motion and epicyclic frequency. The CER force, which is proportional to Galatea's eccentricity, e_G , will be stronger than the CIR, which is proportional to the square of Galatea's inclination, I_G . Both quantities are small: $I_G = 0.0544 \pm 0.0132 \text{ deg}$ with respect to the invariable plane, and $e_G = (0.120 \pm 0.149) \times 10^{-3}$ (ref. 7), though only I_G is measurably non-zero. It is possible to shift the CER exactly at the arcs' current location by accounting for the arcs' inertia, which has the effect of pulling on Galatea's apsidal line through the coupling of the CER to the Lindblad resonances LR1 and LR2 (Table 1). An estimate of the mass of the ring that would be required to switch on the resonance can be obtained from the condition that the CER angle is stationary. Exact resonance requires that the mismatch between the arcs' current mean motion⁵, $n = 820.1122 \text{ deg d}^{-1}$, and the CER pattern speed, n_{CER} , equal the precession rate of Galatea's pericentre induced by the massive ring and denoted $\varpi_{\text{G,ring}}$:

$$\varpi_{\text{G,ring}} = 43(n - n_{\text{CER}}) \quad (1)$$

Three main effects contribute to $\varpi_{\text{G,ring}}$: first, the CER as the strongest of the three induces a regression of Galatea's pericentre given by:

$$\varpi_{\text{G,CER}} = m\alpha M^{-1} f_{\text{CER}}(\alpha) n_G e_G^{-1} \quad (2)$$

when m is the ring's mass, M is Neptune's mass, $\alpha = 0.9844$ is the semi-major axes ratio of Galatea and the arcs, $f_{\text{CER}}(0.9844) = -35.65$ is the CER strength evaluated with the observed semi-major axes, and $n_G = 839.6615 \text{ deg d}^{-1}$ (ref. 5) is Galatea's mean motion. Second, the LR2 also contributes a regression of Galatea's pericentre given by:

$$\varpi_{\text{G,LR2}} = m\alpha M^{-1} f_{\text{LR2}}(\alpha) n_G e_G^{-1} \quad (3)$$

where $f_{\text{LR2}}(0.9844) = -5941.82$ is the LR2 strength. Third, the secular potential that arises from the averaged motion of Galatea and the ring causes a precession of the satellite's pericentre given by:

$$\varpi_{\text{G,sec}} = (2\pi)^{-1} m\alpha M^{-1} n_G (1 - \alpha)^{-2} \quad (4)$$

Substituting equations (2)–(4) into equation (1), the ring's mass is given by:

$$m = \frac{43(n - n_{\text{CER}}) M e_G}{n_G \alpha (f_{\text{CER}} + f_{\text{LR2}} e + (2\pi)^{-1} (1 - \alpha)^{-2} e_G)} \quad (5)$$

This relation implies that for a given ring mass and satellite eccentricity, arcs can be stabilized in resonance over a continuum of semi-major axes (Fig. 2). The current uncertainty in Galatea's eccentricity allows a wide range of masses for the ring system: from $0.23 m_G$ where $m_G = 2.1 \times 10^{21} \text{ g}$ is Galatea's mass for $e_G = 10^{-4}$, to $0.002 m_G$ for $e_G = 10^{-6}$. However, as the resonance width, W_{CER} , is related to e_G through $W_{\text{CER}} = 2.5(e_G/10^{-4})^{1/2} \text{ km}$, the arcs' current semi-major axis width, W_{arc} , can provide a constraint on Galatea's eccentricity and therefore the ring's mass by equating W_{arc} and W_{CER} . Unfortunately, the arcs' true spread in semi-major axis is poorly known. We choose the nominal value $W_{\text{arc}} = 0.4 \text{ km}$ corresponding to that of the CIR resonance model and the corresponding kinematic fit². This yields $e_G = 2.5 \times 10^{-6}$ and a ring mass of $0.006 m_G$. In Fig. 1e, f, a particle of mass $0.002 m_G$ is shown to librate in the 43:42 CER for $e_G = 10^{-6}$ —a detailed analysis (F.N. and C.P., to

Table 1 Galatea's resonances

Resonance type	Resonance angle	Δa (km)
Eccentric corotation	$43\lambda - 42\lambda_G - \varpi_G$	-1.99
Lindblad 2	$86\lambda - 84\lambda_G - \varpi_G - \varpi$	-1.97
Lindblad 1	$43\lambda - 42\lambda_G - \varpi$	-1.94
Parametric	$86\lambda - 84\lambda_G - 2\Omega$	-0.33
Vertical	$86\lambda - 84\lambda_G - \Omega_G - \Omega$	-0.31
Inclined corotation	$86\lambda - 84\lambda_G - 2\Omega_G$	-0.29

Shown are the inventory of resonances associated with Galatea in the arcs' neighbourhood. The resonance distance to the arcs, Δa , is referred to the new location⁵ $62,932.85 \text{ km}$ and has been shifted by $+0.15 \text{ km}$ to account for the synodic change in mean motion that is due to Galatea's action on the ring-particle¹⁴. The variables λ , ϖ , Ω , λ_G , ϖ_G , Ω_G are respectively the mean longitude, the longitude of the pericentre and the longitude of the ascending node of a ring-particle and Galatea.

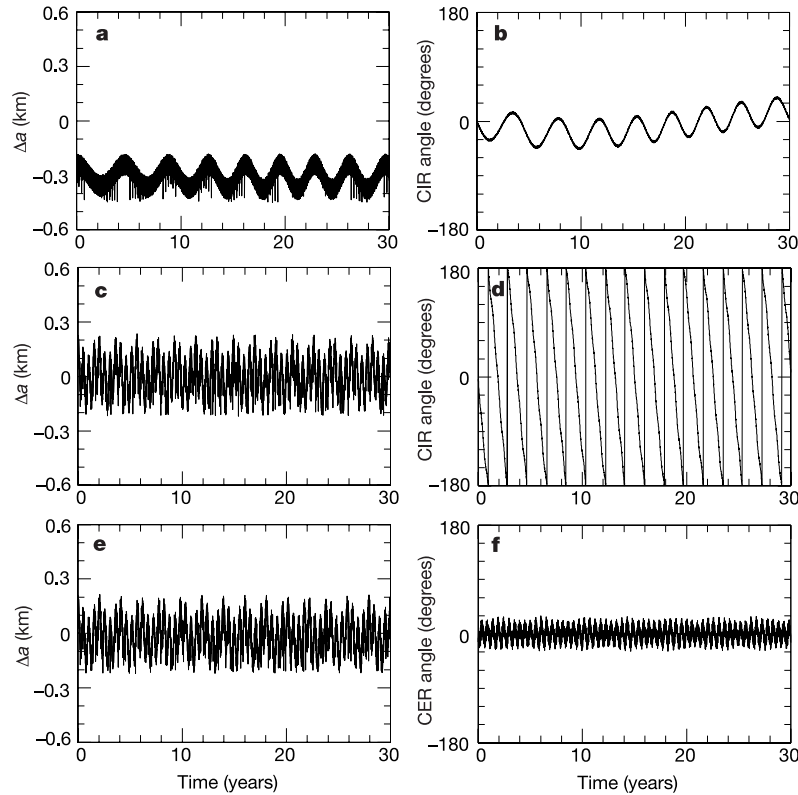


Figure 1 The motions of arc particles around Neptune. Numerical integration of the motion of an arc particle interacting with Galatea (mass 2.1×10^{21} g, eccentricity 10^{-6} , inclination 0.052 deg, mean motion $839.6615 \text{ deg d}^{-1}$, corresponding to a geometric semi-major axis of 61,952.60 km). **a, c, e**, Δa denotes the geometric semi-major axis of the arc particle referred to the location given by the latest set of observations according to ref. 5, 62,932.85 km. **b, d, f**, The corotation inclination and eccentricity resonance (CIR and CER) angles are respectively $86\lambda - 84\lambda_G - 2\Omega_G - 180$ deg and $43\lambda - 42\lambda_G - \varpi_G$. **a, b**, The old model of angular confinement of the CIR of width 0.45 km. The arcs were thought to be 0.3 km closer to Galatea than the latest observations indicate. The libration of the CIR angle ensures that the collective motion of

the arc particles has the appearance of an arc centred on the corotation equilibrium site. **c, d**, The revised evolution of the arcs after the new observations put the arcs 0.3 km outside the location of the CIR. In this case the CIR angle is circulating through 360 deg and angular confinement is lost. **e, f**, The evolution of a massive arc system with mass ratio to Galatea of 0.002 and located at the arcs' current position. In this case the CER angle librates around 0 deg, ensuring confinement. In addition, the vertical resonance that is almost coincident with the CIR (Table 1) forces a small inclination, 0.019 deg, on the arcs with respect to Neptune's invariable plane¹⁵. This value differs from the provisional Voyager value of 0.062 deg which is more likely to be due to errors in identifying the orientation of Neptune's pole^{2,6}.

be published elsewhere) shows that the previous formulae overestimate the arcs' mass. Decreasing the arcs' mass below the threshold value results in the increase of the CER libration amplitude until the arc particle is no longer librating. Unlike the CIR, the CER can be shifted easily to the arcs' position because the pericentre regression rate it forces on Galatea is proportional to e_G^{-1} which can be large even for a small mass ring. The CIR-induced nodal regression rate on the other hand scales as $H_G^{-1} \approx 1$ resulting in a negligible shift of the resonance location.

The angular confinement by the CER can now explain the angular length of the arc Fraternité⁸, approximately 10 deg, as the result of occupying a single corotation site. This feature remained unexplained in the CIR model because the arc was found to extend over 2 to 3 consecutive corotation sites of width 4.18 deg despite the presence of unstable equilibria within it. This observation, together with the inability of the CIR to be shifted substantially from its nominal location, raises the question: is the CIR effect needed at all to explain the arc structure? The answer lies in the spacings of some of the smaller arcs⁸: the location of arc Égalité 2, of length about 3 deg, 10 deg away from the centre of Fraternité, and 3 deg away from the centre of Égalité 1, of length about 1 deg, does not match the 8.37-deg spacings of the CER. However, the coupling of the CIR and CER potentials at the current location modifies the shape of the CER potential by introducing additional equilibria with smaller

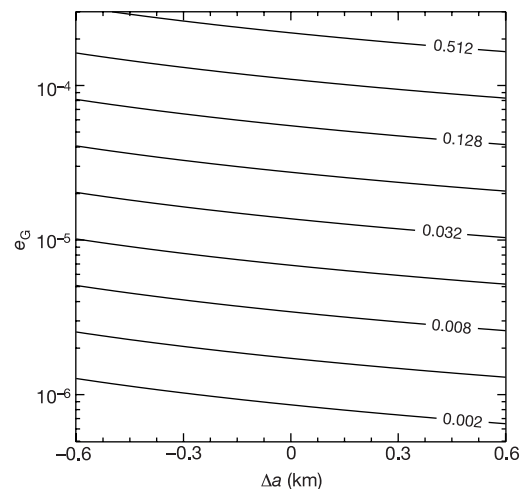


Figure 2 The effect of ring mass on the resonance location. The ring's mass required to shift the 43:42 corotation eccentricity resonance (CER), scaled to Galatea's mass 2.1×10^{21} g, is given by the level curves of equation (5) as a function of the semi-major axis referred to the current position of the arcs Δa , and Galatea's eccentricity e_G . As equation (5) is approximately linear as a function of the eccentricity, a small arc mass is sufficient to shift the CER.

spacings that match the observation of the smaller arcs. Conversely, we note that if the arcs' current mean motion did not differ from that of the CIR, that is, if there were no observed offset, the CER is still the only resonance that can both confine the arcs and explain the longer radial structures; the required mass for shifting the CER to the CIR location can be found in Fig. 2.

The ring mass we determined, $0.002m_G$ assuming $e_G = 10^{-6}$, corresponds to a parent 10-km-radius satellite with a density similar to Galatea's. We note that many of our results can be reproduced in the case of a massless arc system if a second satellite shares its orbit. A similar model was put forward shortly after the discovery of the ring arc system but prior to Voyager's more detailed observations¹⁶. A closer inspection of this two-satellite model however shows that the second satellite would produce a single continuous arc at each equilibrium point (L_4 or L_5), leaving the sequence of small arcs unexplained. A satellite co-orbital to the arcs of mass $0.002m_G$ would activate the CER, which in turn modifies the satellite's potential near L_4 and L_5 to allow a structure similar to the arcs. However, Voyager data exclude⁹ undetected satellites of radius larger than 6 km implying that the mass required for the angular confinement of the arcs is not contained in a single body. Modelling the breakup of an arc parent satellite^{10,11} should henceforth include the dynamical effects described in this paper in order to ascertain the angular distribution of mass and solve the problem of radial stability of the arcs. If the torques exerted by Galatea on the arcs work like those in Saturn's rings^{12,13}, the rate of radial migration of the arcs is 2.4 km yr^{-1} , disrupting the arcs in less than a year. However, the satellite's torque is less likely to cause a significant drift in the arcs' radial position if the rings' mass is concentrated in a few clumps.

The value estimated here for Galatea's eccentricity, about 10^{-6} , is consistent with the requirement that the arcs have a small mass, regardless of the physical model responsible for their angular confinement. The fact that the current spread in semi-major axis of the arcs leads to such a small eccentricity is encouraging, because we expect that the decay timescale of Galatea's eccentricity due to the tides raised by Neptune¹ is of the order of 10^8 years, implying a rapid circularization of the orbit over the age of the Solar System. More precise measurements of Galatea's orbital elements, as well as models of the tidal evolution of the inner neptunian satellites together with the ring-Galatea interaction¹³, are needed to fully validate the resonance model presented here and determine the origin of the small residual eccentricity in Galatea's orbit responsible for the arcs' confinement. □

Received 26 November 2001; accepted 28 February 2002.

1. Goldreich, P., Tremaine, S. & Borderies, N. Towards a theory for Neptune's arc rings. *Astron. J.* **92**, 490–494 (1986).
2. Porco, C. An explanation for Neptune's ring arcs. *Science* **400**, 995–1001 (1991).
3. Dumas, C. *et al.* Stability of Neptune's arcs in question. *Nature* **400**, 733–735 (1999).
4. Sicardy, B. *et al.* Images of Neptune's ring arcs obtained by a ground-based telescope. *Nature* **400**, 731–733 (1999).
5. Dumas, C. *et al.* Astrometry and near infra-red photometry of Neptune's inner satellites and ring arcs. *Astron. J.* (in the press).
6. Nicholson, P. D., Mosquera, I. & Matthews, K. Stellar occultation observations of Neptune's rings: 1984–1988. *Icarus* **113**, 295–330 (1995).
7. Owen, W. M., Vaughan, R. M. & Synnott, S. P. Orbits of the six new satellites of Neptune. *Astron. J.* **101**, 1511–1515 (1991).
8. Porco, C. *et al.* in *Neptune and Triton* (ed. Cruikshank, D. P.) 703–804 (Univ. Arizona Press, Tucson, 1995).
9. Smith, B. A. *et al.* Voyager 2 and Neptune: Imaging science results. *Science* **246**, 1422–1449 (1989).
10. Colwell, J. E. & Esposito, L. W. A model for dust production for Neptune's ring system. *Geophys. Res. Lett.* **17**, 1741–1744 (1990).
11. Colwell, J. E. & Esposito, L. W. Origin of the rings of Neptune and Uranus: II Initial conditions and ring moon population. *J. Geophys. Res.* **98**, 7387–7401 (1993).
12. Goldreich, P. & Tremaine, S. The dynamics of planetary rings. *Annu. Rev. Astron. Astrophys.* **20**, 249–283 (1982).
13. Namouni, F. Ringlet-satellite interactions. *Mon. Not. R. Astron. Soc.* **300**, 915–930 (1998).
14. Horanyi, M. & Porco, C. Where exactly are the arcs of Neptune? *Icarus* **106**, 525–535 (1993).
15. Hänninen, J. & Porco, C. Collisional simulations of Neptune's ring arcs. *Icarus* **126**, 1–27 (1997).
16. Lissauer, J. J. Shepherding model for Neptune's arc ring. *Nature* **318**, 544–545 (1985).

Acknowledgements

We thank C. Dumas and P. Nicholson for reviews and B. Bottke, R. Canup, M. Evans, P. Goldreich, D. Hamilton, D. Nesvorný, J. Peterson, H. Throop and B. Ward for discussions. We acknowledge support from Southwest Research Institute's Internal Research Grant programme and NASA's Planetary Geology and Geophysics Discipline programme.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.N. (e-mail: fathi@ciclops.swri.edu).

Towards Bose–Einstein condensation of excitons in potential traps

L. V. Butov*, C. W. Lai*, A. L. Ivanov†, A. C. Gossard‡ & D. S. Chemla*§

* Materials Sciences Division, E. O. Lawrence Berkeley National Laboratory;

§ Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA

† Department of Physics and Astronomy, Cardiff University, Cardiff CF24 3YB, UK

‡ Department of Electrical and Computer Engineering, University of California, Santa Barbara, California 93106, USA

An exciton is an electron–hole bound pair in a semiconductor. In the low-density limit, it is a composite Bose quasi-particle, akin to the hydrogen atom¹. Just as in dilute atomic gases^{2,3}, reducing the temperature or increasing the exciton density increases the occupation numbers of the low-energy states leading to quantum degeneracy and eventually to Bose–Einstein condensation (BEC)¹. Because the exciton mass is small—even smaller than the free electron mass—exciton BEC should occur at temperatures of about 1 K, many orders of magnitude higher than for atoms. However, it is in practice difficult to reach BEC conditions, as the temperature of excitons can considerably exceed that of the semiconductor lattice. The search for exciton BEC has concentrated on long-lived excitons: the exciton lifetime against electron–hole recombination therefore should exceed the characteristic timescale for the cooling of initially hot photo-generated excitons^{4–10}. Until now, all experiments on atom condensation were performed on atomic gases confined in the potential traps. Inspired by these experiments, and using specially designed semiconductor nanostructures, we have collected quasi-two-dimensional excitons in an in-plane potential trap. Our photoluminescence measurements show that the quasi-two-dimensional excitons indeed condense at the bottom of the traps, giving rise to a statistically degenerate Bose gas.

More than three decades ago Keldysh and Kozlov¹ showed that in the dilute limit, $na_B^D \ll 1$ (a_B is the exciton Bohr radius, n the exciton density and D the dimensionality), excitons behave as weakly interacting Bose particles and are expected to undergo the Bose–Einstein condensation (BEC). Because the exciton mass, M , is very small, the three-dimensional critical temperature for exciton BEC, $T_c^{3D} = 0.527k_B^{-1}2\pi\hbar^2M^{-1}(n/g)^{2/3}$ (g is the spin degeneracy of the exciton state, k_B is the Boltzmann constant), should reach several Kelvins at experimentally accessible exciton densities, that is, about six orders of magnitude higher than the critical temperature for atom BEC. The theoretical predictions for exciton BEC and

a possible crossover from exciton BEC at low densities to the Bardeen–Cooper–Schrieffer (BCS)-like condensate, the excitonic insulator, at high densities ($na_B^D \gg 1$) (ref. 11) have initiated intensive experimental studies^{4,5,7,8,9–10}.

In spite of the relatively high T_c , it is experimentally very difficult to lower the temperature of a high-density exciton gas enough to reach the BEC. Cooling semiconductor lattices well below 1 K is routinely achieved with He refrigerators. Nevertheless, the exciton temperature, T_x , can often considerably exceed that of the lattice. In fact, because the excitons can recombine, T_x is determined by the ratio of the exciton energy relaxation and recombination rates. Therefore, the search for BEC has concentrated on long-lived excitons: to reach low enough T_x after the initial photo-generation, the exciton gas should be given time to lower its temperature to close to that of the lattice.

Over the last two decades the experimental efforts to observe exciton BEC in bulk semiconductors dealt mainly with bulk Cu_2O (refs 4, 5), a material whose ground exciton state is optically dipole-inactive and has, therefore, a very low radiative recombination rate. The conditions necessary for realization of BEC in Cu_2O appeared to be experimentally accessible: $T_c \propto n^{2/3}$ would reach ≈ 2.3 K at exciton density $n = 10^{17} \text{ cm}^{-3}$. However, another, competitive density-dependent process begins at high n : the exciton Auger recombination rate, proportional to n , which increases faster than T_c with increasing n . Recent precise measurements indicate that the Auger recombination rate in Cu_2O is about two orders of magnitude higher than was assumed before and, therefore, the exciton densities reached in the up-to-date experiments are far below that

required to achieve a degenerate Bose gas of excitons⁹.

In our attempt to create a degenerate Bose gas of excitons we explore another system: a quasi-two-dimensional gas of indirect excitons in coupled quantum wells (QWs)^{7,10}. There are two primary advantages of using coupled QWs for realizing a quantum degenerate exciton gas and, ultimately, the BEC-like phase transition: (1) the long lifetime of indirect excitons and (2) the short exciton cooling time. The long lifetimes originate from the spatial separation between the electron and hole layers (Fig. 1a), resulting in a radiative lifetime of indirect excitons in our coupled QW samples that is more than three orders of magnitude longer than that of direct excitons in single QWs. As for the cooling of hot photoexcited excitons down to the temperatures of the cold lattice, which occurs via emission of bulk longitudinal acoustic phonons, it is about three orders of magnitude faster for excitons in GaAs QWs than that in bulk GaAs. This is due to relaxation of the momentum conservation law in the z direction ($z \perp \text{QW plane}$). Indeed, for quasi-two-dimensional systems the ground-state mode $E = 0$ couples to the continuum of the energy states $E \geq E_0$ rather than to the single energy state $E = E_0 = 2Mv_s^2$ (v_s is the sound velocity) as it occurs in bulk semiconductors⁸. The long lifetime and the fast cooling rate of indirect excitons favour accumulation of these Bose particles in the lowest energy states and allow the photoexcited excitons to cool down to temperatures well below 1 K, where the dilute quasi-two-dimensional gas of indirect excitons becomes statistically degenerate¹⁰.

Coupled QW structures can provide a further advantage for indirect-exciton BEC. It is worth noting here that all the recent

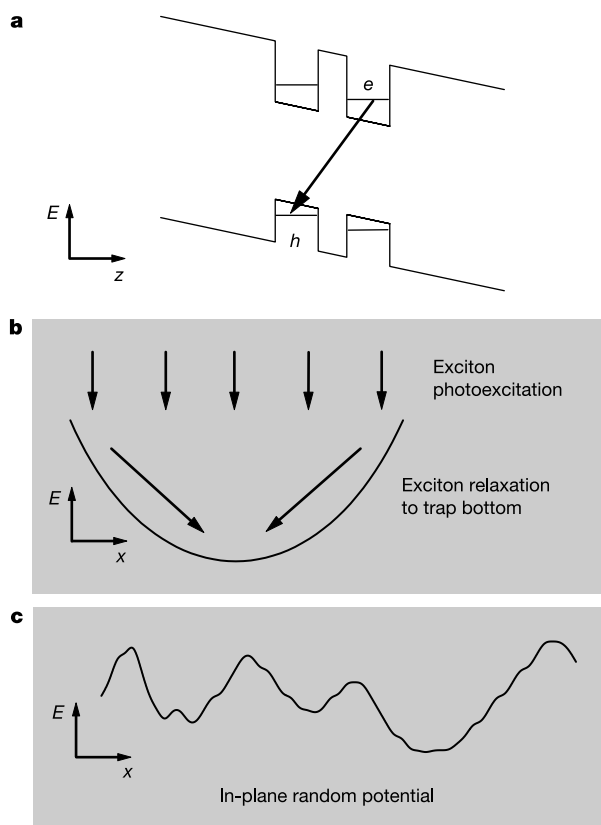


Figure 1 Potential profiles in coupled quantum well structures in the growth direction, z , and in the x – y plane. **a**, Energy band diagram of the coupled quantum well (QW) structures. The indirect excitons in coupled QWs are characterized by high cooling rates, three orders of magnitude higher than in bulk GaAs, and long lifetime, more than three orders of magnitude longer than in a single GaAs QW. The lifetime is much longer than the characteristic timescale for cooling, so that the exciton temperature can be very low, well

below 1 K. **b**, In-plane potential of a trap for excitons in QW structures. The conditions for the exciton BEC are significantly improved in such potential traps because of the confinement effect and the enhanced local density driven by the exciton energy relaxation. **c**, Schematic of the in-plane potential landscape formed by random potential fluctuations. Such natural local potential minima can host highly degenerate exciton systems and, ultimately, BEC.

advances in BEC of atoms were made possible by exploiting the confinement of atomic gases within potential traps. Similarly, the conditions for BEC of indirect excitons are significantly improved if the excitons are collected in a local minimum caused by an in-plane modulation of the potential (Fig. 1b). Then T_c increases because of confinement effects and the enhancement of the local density at the trap centre that result from the drift of photoexcited excitons toward the bottom of the trap.

In particular, in the case of confinement by an in-plane two-dimensional harmonic potential the critical temperature is:

$$T_c = \frac{1}{k_B} \frac{\sqrt{6}}{\pi} \hbar \omega \sqrt{N/g} \quad (1)$$

where $\hbar \omega$ is the quantization energy and N is the number of excitons in the trap. In the case of confinement by an in-plane square potential (two-dimensional box):

$$T_c = \frac{4\pi \hbar^2 n}{2Mgk_B \ln(nS/g)} \quad (2)$$

where S is the area of the box¹². We note that the T_c given by these two equations vanishes either when $\hbar \omega \rightarrow 0$ or when $S \rightarrow \infty$, thus reproducing the well-established results that in infinite ideal two-

dimensional systems BEC is only possible at $T = 0$. In such a system, however, a phase transition to a superfluid exciton state is still possible at finite temperatures¹³. As for atomic systems, excitons condense in real space in the lowest energy state at the bottom of the trap, which corresponds to the condensation in momentum space for infinite systems, that is, to the BEC.

Semiconductor technology allows us now to engineer the in-plane potential for indirect excitons in nanostructures. For example, traps can be formed by lateral modulation of the electric field along z , E_z , by application of a local stress or of a local magnetic field. Besides this interesting approach intrinsic local minima of the in-plane potential are always present in any QW structure. These natural traps are caused by random fluctuations of the interfaces, for example, QW thickness and/or alloy disorder, or E_z (see Fig. 1c). Such potential minima can host highly degenerate exciton systems and ultimately BEC^{6,7}. There is a great variety of fluctuations (such as in width and depth) in the random potential of any semiconductor heterostructure. Therefore, among all the many traps present in a given sample, the exciton gas will, by itself, select those with optimal parameters for condensation: this is an important advantage of the natural traps. A clear signature of condensation of a degenerate exciton gas in such a trap is its strongly in-plane localized photoluminescence (PL) emission peaked at the bottom

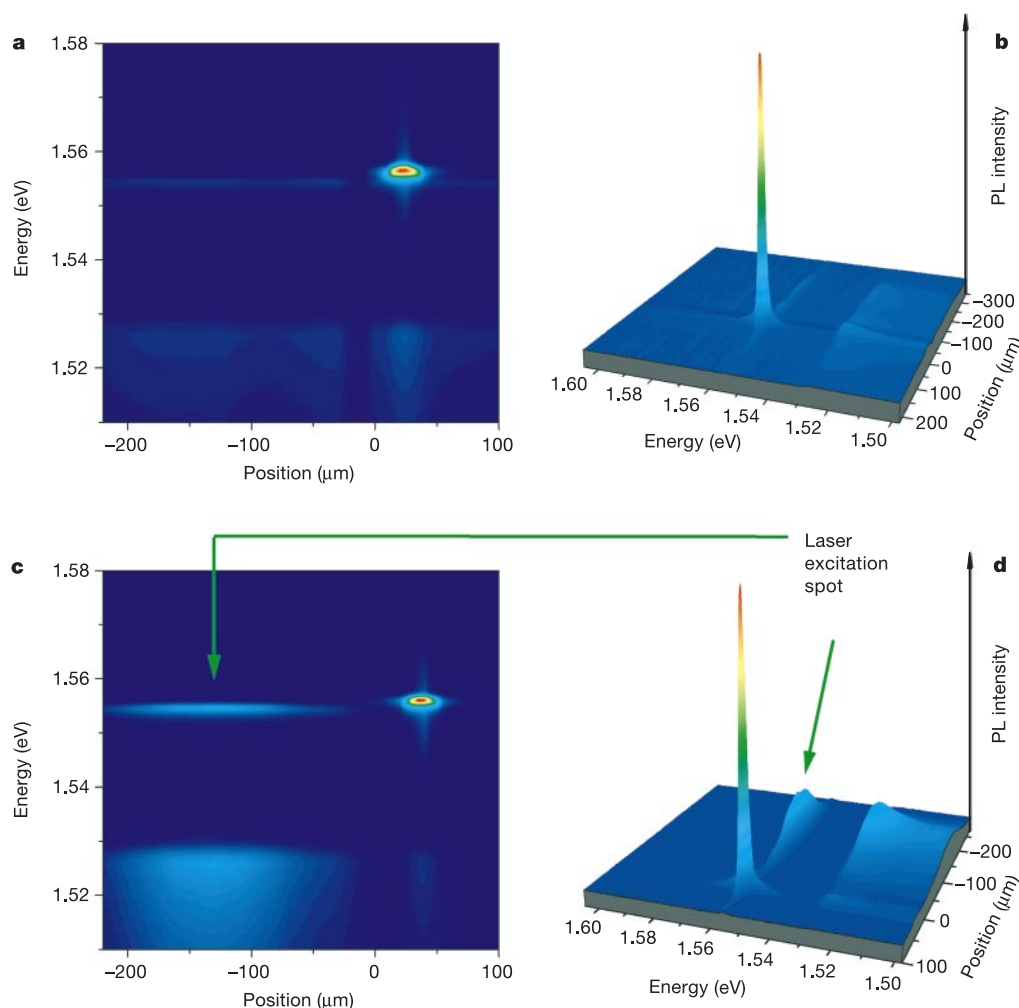


Figure 2 Photoluminescence intensity versus energy and one of the in-plane coordinates x at $T = 1.6$ K (the photoluminescence (PL) variation along the orthogonal in-plane coordinate y is the same). The indirect exciton PL line is at about 1.555 eV; the broad line below 1.53 eV comes from the bulk n^+ -GaAs emission. **a, b**, The laser excitation spot is defocused and excites the whole $500 \times 500 \mu\text{m}^2$ mesa. The indirect excitons are

collected in a natural trap resulting in a PL intensity at its centre about 30 times higher than outside. **c, d**, The laser excitation is focused to a spot around $100 \times 200 \mu\text{m}^2$ about $170 \mu\text{m}$ away from the trap. Even in these conditions the indirect exciton PL intensity at the trap centre is about six times larger than at the centre of the excitation spot. This demonstrates drift and condensation of indirect excitons over macroscopic distances.

of the trap.

Here we report the first observation of the condensation of indirect excitons in natural traps, where a highly statistically degenerate Bose gas of indirect excitons builds up. We have performed spectrally and spatially resolved PL, monitoring both the indirect and direct exciton emissions. This enables us to compare, under identical experimental conditions, the anomalous behaviour of the former and the regular behaviour of the latter (for details see Methods section). Using a spatial resolution of $10\text{ }\mu\text{m}$ at $T = 1.6\text{ K}$, we have observed a huge local enhancement of the indirect exciton PL intensity in several $500 \times 500\text{ }\mu\text{m}^2$ mesas. We interpret this effect by the effective accumulation of excitons in specific natural traps. Depending on the location on the wafer, the mesas can exhibit no trap, one trap or two traps. In a given sample the position of traps is well-defined and all experimental results are extremely robust: for example, they are reproducible after cycling the sample temperature up to room temperature and back to 1.6 K many times. Study of the spatial dependence of the PL intensity along two orthogonal in-plane directions, x and y , indicates that the exciton clouds in the traps have a nearly circular shape.

As shown in Fig. 2a and b, under uniform excitation of a whole mesa using a defocused laser, the indirect exciton PL intensity at the centre of a trap is about 30 times higher than that emitted from any other location on the mesa outside the trap. This clearly demonstrates that the indirect excitons collect at the centre of the trap, where their concentration is strongly enhanced. We estimate the density of indirect excitons in the uniformly illuminated mesas, far from the traps, to be $n_{\text{avg}} \approx 3 \times 10^9\text{ cm}^{-2}$. This value, using the PL

intensities ratio for the exciton density in the traps yields: $n_{\text{trap}} \approx 10^{11}\text{ cm}^{-2}$. In Fig. 2a and b the full width at half maximum (FWHM) of the profile of the exciton cloud in the trap is $23\text{ }\mu\text{m}$; thus, the total number of excitons it contains is $N = nS \approx 4 \times 10^5$.

We have also explored the traps under inhomogeneous excitation using a smaller laser spot, $100 \times 200\text{ }\mu\text{m}^2$, that can be located anywhere on the mesa as far away from a trap as $400\text{ }\mu\text{m}$. As seen in Fig. 2c and d, even in that case the indirect excitons are collected by the trap and the PL intensity from the trap is up to about six times higher than the PL intensity in the excitation spot. The long exciton lifetime and large diffusion coefficient allow the indirect excitons to travel several hundred micrometres and to be collected by the trap. During their travel the initially hot excitons, photo-excited at high energy, effectively thermalize down to the lattice temperature via photon emission. Therefore, in the case of remote excitation the trap is filled by the cold excitons. This is confirmed by the disappearance of the PL emission line of the hot direct excitons from the spectrum in the trap, although that line is present in the spectra at the excitation spot. In this case the exciton density in the laser spot is $n_{\text{spot}} \approx 2 \times 10^{10}\text{ cm}^{-2}$, which implies $n_{\text{trap}} \approx 1.2 \times 10^{11}\text{ cm}^{-2}$ in the trap. This value is similar to that obtained in experiments with defocused laser excitation discussed above. In Fig. 2c and d the FWHM of the profile of the exciton cloud in the trap is $20\text{ }\mu\text{m}$ so that $N = nS \approx 4 \times 10^5$.

Assuming that one can apply the thermodynamic equilibrium description of ideal point-like bosons to indirect excitons and a two-dimensional box shape for the trap, equation (2) with the GaAs parameters $g = 4$ and $M = 0.21m_0$ (m_0 is the free electron mass¹⁴)

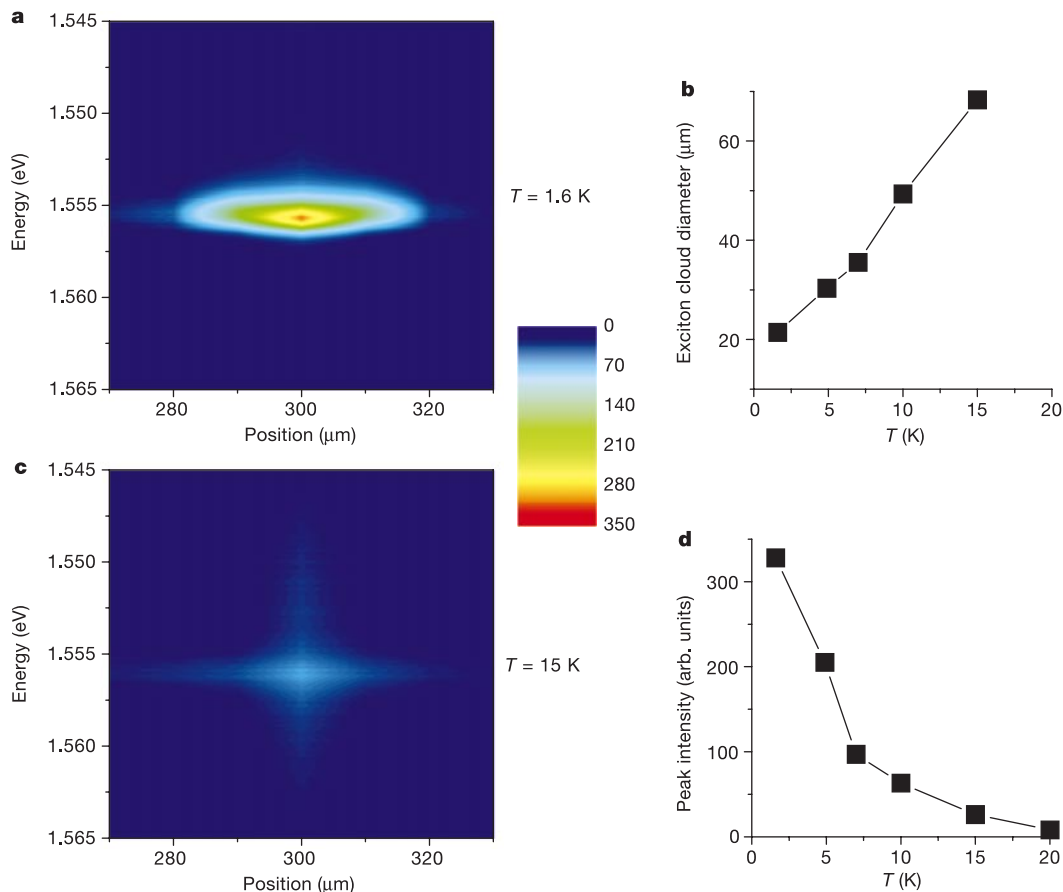


Figure 3 The spatial shrinkage of the exciton cloud near the bottom of the potential trap as the temperature is reduced. **a, c**, Contour map of the indirect exciton PL intensity versus the in-plane coordinate x and energy at $T = 1.6\text{ K}$ (**a**) and $T = 15\text{ K}$ (**c**). The temperature

dependence of the diameter of the indirect exciton cloud and their PL peak intensity are shown in **b** and **d**. As the temperature is reduced the indirect excitons condense at the bottom of the trap.

gives an estimate for the critical temperatures for exciton BEC in the trap of $T_c \approx 0.5$ K and $T_c \approx 0.7$ K, for the uniform and remote excitations respectively. These crude estimates imply that in our current experiments with $T_{\text{bath}} = 1.6$ K the conditions for exciton BEC are still not achieved. We note, however, that many factors are disregarded in the estimate; in particular, the indirect excitons are not ideal bosons—on the contrary, they are characterized by a nonlocal repulsive interaction that reinforces condensation and should result in an increase of T_c (ref. 15). In the remote excitation experiments the degenerate Bose gas of indirect excitons builds up outside the trap in the excitation spot, as revealed by the bosonic stimulation of exciton scattering to the low-energy states (similar to experiments in ref. 10). The degree of quantum degeneracy is much higher for indirect excitons accumulated at the centre of the trap.

The diameter of the exciton cloud near the bottom of the potential trap reduces as the temperature is lowered, while the peak intensity increases, as shown in Fig. 3. This demonstrates that while at high temperatures the excitons are distributed over the high energy states of the trap, at low temperatures the excitons condense at the bottom of the trap. Similar spatial shrinkage of atom clouds is characteristic of atomic BEC in the potential traps.

Direct excitons, however, exhibit none of the collection effects which influence the indirect excitons so strongly. This is consistent with a high temperature of direct excitons: the short lifetime does not allow an effective cooling of direct excitons. Furthermore, that short lifetime limits the distance they can travel before annihilation, making collection effects on large length scales impossible for direct excitons.

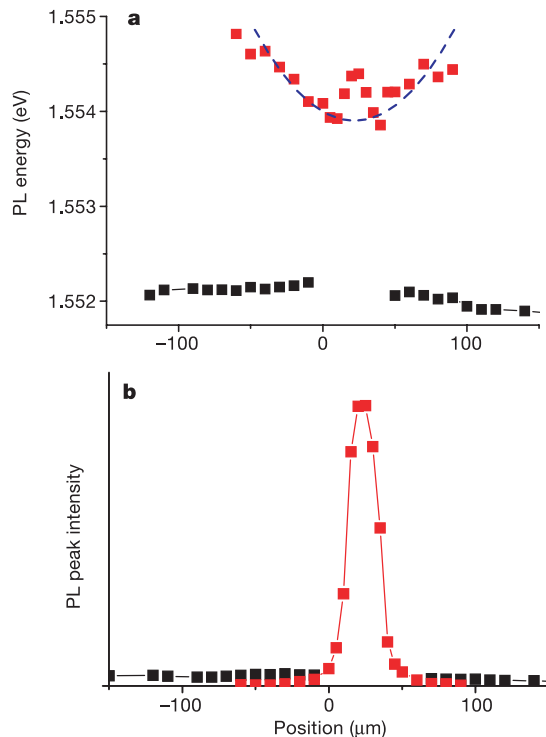


Figure 4 Anomalous properties of the traps revealed by the spatial dependence of the indirect exciton photoluminescence. The energy of the PL peak is higher in the trap than outside; although at large distances from the centre, $\geq 10 \mu\text{m}$, the PL energy indicates that the potential in the trap has an overall bowl-like shape, it exhibits a 0.5-meV bump near the centre that increases with density, $T = 1.6$ K. **a**, **b**, Red (black) squares present the indirect exciton emission energies (**a**) and intensities (**b**) in (out) of the potential trap. Both emission lines are resolved near the trap edge. The blue dashed line is a guide for the eyes and represents a parabolic profile.

In Fig. 4 we show how the PL of optically active indirect excitons varies with the distance ρ from the centre of the trap. The PL emission energy exhibits several anomalies. First, the energy is higher inside the trap than outside by $\Delta E_{\text{PL}} \approx 2$ meV. This difference remains practically constant when the indirect exciton energy inside and outside the trap are reduced by increasing E_z : the trap looks like a 'levitating bowl'. This raises several interesting issues for future studies: (1) Although ΔE_{PL} is much smaller than the initial energy of the photoexcited indirect excitons, how can the trap collect these indirect excitons despite a non-negligible energy barrier at its edges? (2) The indirect exciton energy in the trap can be limited by its depth in analogy with the case of atom evaporative cooling in optical traps. Second, the PL emission energy exhibits a bump at the centre of the trap, $\delta E_{\text{PL}} \approx 0.5$ meV, which increases with the density of indirect excitons. We tentatively explain such a 'Mexican hat' structure by the pressure induced by the stimulated flux of excitons towards the centre of the trap. The microscopic origin of the trap is not essential. The only important point is that in the fluctuating in-plane potential landscape of the sample the exciton gas finds spontaneously those local minima with optimal parameters for condensation. The possibility of evaporative cooling could be important in the potential minimum selection. Modelling the overall trap profile by the harmonic potential shown by the blue dashed line in Fig. 4, equation (1) with $N = 4 \times 10^5$ gives, for the 'ideal boson' case, $T_c \approx 1.0$ K, close to the two-dimensional box value. □

Methods

n^+i-n^+ GaAs/AlGaAs coupled QW structure was grown by MBE. The i -region consists of two 8-nm GaAs QWs separated by a 4-nm $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ barrier and surrounded by two 200-nm $\text{Al}_{0.33}\text{Ga}_{0.67}\text{As}$ barrier layers. The electric field in the z -direction is monitored by the external gate voltage V_g applied between the highly conducting n^+ -layers. The narrow PL linewidth, about 1 meV, indicates small in-plane disorder. The sample was excited with a HeNe laser, wavelength $\lambda = 632.8$ nm and power $P = 3$ mW. The experiments were performed in a ^4He cryostat with optical windows. A spatial resolution of $10 \mu\text{m}$ was achieved using a pinhole in the intermediate image plane. For $V_g \approx 0$ the ground state of the optically pumped coupled QW is a direct exciton made of an electron and a hole in the same layer and similar to the excitons in single QWs. For nonzero V_g the ground state is an indirect exciton made of an electron and a hole in different layers (Fig. 1a). The indirect excitons have a permanent dipole along z and thus exercise on each other a repulsive interaction—their energy therefore increases with density. The corresponding energy shift, $\delta E(n)$, allows us to evaluate their concentration using $\delta E(n) = 4\pi n e^2 d / \epsilon$, where d is the effective separation between the electron and hole layers and ϵ is the dielectric constant. For the strongest excitations in the experiments reported here excitation density $W = 10 \text{ W cm}^{-2}$ and maximum density $n_{\text{max}} \approx 2 \times 10^{10} \text{ cm}^{-2}$ at the centre of the excitation spot. Radiative recombination is the dominant decay mechanism of indirect excitons in our high-quality sample. The spatially resolved PL kinetics shows that the excitons in the centre of the trap have a lifetime as long as the excitons outside the trap. The energy of the trapped excitons varies with applied electric field similar to that of the excitons outside the trap, thus proving unambiguously that the long-lived indirect excitons are those that condense at the bottom of the trap.

Received 26 October 2001; accepted 22 February 2002.

- Keldysh, L. V. & Kozlov, A. N. Collective properties of excitons in semiconductors. *Zh. Eksp. Teor. Fiz.* **54**, 978–993 (1968); *Sov. Phys. JETP* **27**, 521–528 (1968).
- Anderson, M. N., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Observation of Bose–Einstein condensation in a dilute atomic vapor. *Science* **269**, 198–202 (1995).
- Davis, K. B. *et al.* Bose–Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75**, 3969–3973 (1995).
- Snoke, D. W., Wolfe, J. P. & Mysyrowicz, A. Quantum saturation of a Bose gas: excitons in Cu_2O . *Phys. Rev. Lett.* **59**, 827–830 (1987).
- Lin, J. L. & Wolfe, J. P. Bose–Einstein condensation of paraexcitons in stressed Cu_2O . *Phys. Rev. Lett.* **71**, 1222–1225 (1993).
- Zhu, X., Littlewood, P. B., Hybertsen, M. & Rice, T. Exciton condensate in semiconductor quantum well structures. *Phys. Rev. Lett.* **74**, 1633–1636 (1995).
- Butov, L. V. & Filin, A. I. Anomalous transport and luminescence of indirect excitons in AlAs/GaAs coupled quantum wells as evidence for exciton condensation. *Phys. Rev. B* **58**, 1980–2000 (1998).
- Ivanov, A. L., Littlewood, P. B. & Haug, H. Bose–Einstein statistics in thermalization and photoluminescence of quantum well excitons. *Phys. Rev. B* **59**, 5032–5048 (1999).
- O'Hara, K. E., Süllenhain, L. Ö. & Wolfe, J. P. Strong nonradiative recombination of excitons in Cu_2O and its impact on Bose–Einstein statistics. *Phys. Rev. B* **60**, 10565–10568 (1999).
- Butov, L. V. *et al.* Stimulated scattering of indirect excitons in coupled quantum wells: Signature of a degenerate Bose-gas of excitons. *Phys. Rev. Lett.* **86**, 5608–5611 (2001).
- Keldysh, L. V. & Kopayev, Yu. V. Possible instability of the semimetallic state toward Coulomb interaction. *Fizika Tverdogo Tela* **6**, 2791–2798 (1964); *Sov. Phys. Solid State* **6**, 2219–2224 (1965).
- Ketterle, W. & van Druten, N. J. Bose–Einstein condensation of a finite number of particles trapped in one or three dimensions. *Phys. Rev. A* **54**, 656–660 (1996).

13. Popov, V. N. On the theory of the superfluidity of two- and one-dimensional Bose systems. *Theor. Math. Phys.* **11**, 565–573 (1972).
14. Butov, L. V., Mintsev, A. V., Lozovik, Yu. E., Campman, K. L. & Gossard, A. C. From spatially indirect to momentum-space indirect excitation by in-plane magnetic field. *Phys. Rev. B* **62**, 1548–1551 (2000).
15. Leggett, A. J. Bose–Einstein condensation in the alkali gases: Some fundamental concepts. *Rev. Mod. Phys.* **73**, 307–356 (2001).

Acknowledgements

We thank K. L. Campman for growing the high-quality coupled QW samples. This work was supported by the Office of Energy Research, Office of Basic Energy Sciences and the Division of Material Sciences of the US DOE and by the EPSRC and RFBR.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.V.B.
(e-mail: LVButov@lbl.gov).

All-metallic three-dimensional photonic crystals with a large infrared bandgap

J. G. Fleming*, S. Y. Lin*, I. El-Kady†, R. Biswas & K. M. Ho†

* MS 0603, Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA

† Ames Laboratory, Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA

Three-dimensional (3D) metallic crystals are promising photonic bandgap^{1–3} structures: they can possess a large bandgap^{4–6}, new electromagnetic phenomena can be explored^{7–9}, and high-temperature (above 1,000 °C) applications may be possible. However, investigation of their photonic bandgap properties is challenging, especially in the infrared and visible spectrum, as metals are dispersive and absorbing in these regions¹⁰. Studies of metallic photonic crystals have therefore mainly concentrated on microwave and millimetre wavelengths^{8,11,12}. Difficulties in fabricating 3D metallic crystals present another challenge, although emerging techniques such as self-assembly^{13,14} may help to resolve these problems. Here we report measurements and simulations of a 3D tungsten crystal that has a large photonic bandgap at infrared wavelengths (from about 8 to 20 μm). A very strong attenuation exists in the bandgap, ~ 30 dB per unit cell at 12 μm . These structures also possess other interesting optical properties; a sharp absorption peak is present at the photonic band edge, and a surprisingly large transmission is observed in the allowed band, below 6 μm . We propose that these 3D metallic photonic crystals can be used to integrate various photonic transport phenomena, allowing applications in thermophotovoltaics and blackbody emission.

The tungsten 3D photonic crystal was made by selectively removing Si from already fabricated polysilicon/SiO₂ structures, and back-filling the resulting mould with chemical vapour deposited (CVD) tungsten. This method could be extended to create almost any 3D single-crystal metallic photonic crystal at infrared wavelengths; such photonic crystals have not previously been achievable. Scanning electron microscope (SEM) images of the fabricated four-layer 3D tungsten photonic crystal are shown in Fig. 1a and b.

The optical properties of the 3D tungsten photonic crystal were

characterized using a Fourier-transform infrared measurement system for wavelengths λ ranging from 1.5 to 25 μm (ref. 15). To obtain reflectance (R), a sample spectrum was taken from a 3D tungsten crystal, and then normalized to a reference spectrum of a uniform silver mirror. To find the absolute transmittance (T), a transmission spectrum taken from a 3D tungsten crystal sample was normalized to that of a bare silicon wafer. This normalization procedure is intended to calibrate away extrinsic effects, such as light reflection at the air–silicon interface and silicon absorption. For tilt-angle transmission measurements, the sample was mounted onto a rotational stage with rotational angles θ ranging from 0° to 60°, measured from the surface normal—that is, the $\langle 001 \rangle$ direction.

In Fig. 2a we show the absolute reflectance (black diamonds), transmittance (small black circles) and absorptance (red circles) of a four-layer 3D tungsten photonic crystal. Light propagates along the $\langle 001 \rangle$ direction of the crystal and is unpolarized. The reflectance exhibits oscillations at $\lambda < 5.5 \mu\text{m}$, rises sharply at $\lambda \approx 6 \mu\text{m}$ (the band edge) and finally reaches a high reflectance of 90% for $\lambda > 8 \mu\text{m}$. Correspondingly, the transmittance shows distinct peaks at $\lambda < 5.5 \mu\text{m}$, decreases sharply at $\lambda \approx 6 \mu\text{m}$ (the photonic band edge) and then decreases to below 1% for $\lambda > 8 \mu\text{m}$. The dashed line in Fig. 2a is for reference purposes, and shows the transmittance of a 6,000-Å uniform tungsten film. The simultaneous high R and low T at $\lambda > 8 \mu\text{m}$ is indicative of the existence of a photonic bandgap in the tungsten 3D photonic crystal. The attenuation is as large as ~ 30 dB at $\lambda = 10 \mu\text{m}$ for our four-layer sample, or equivalently a unit cell. The multiple oscillations at $\lambda < 5.5 \mu\text{m}$ are attributed to photonic density-of-states (DOS)

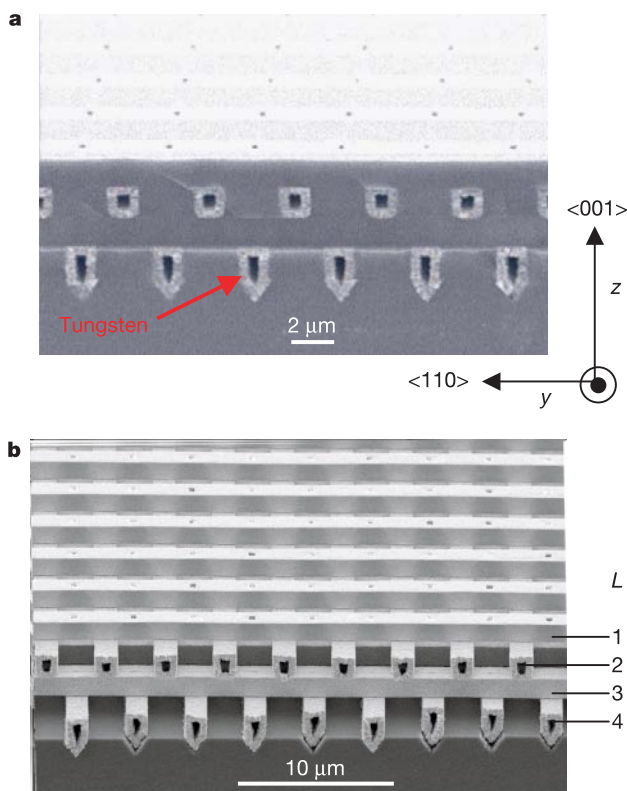


Figure 1 Images of a 3D tungsten photonic crystal, taken by a scanning electron microscope (SEM). The images taken with and without oxide are shown in **a** and **b**, respectively. The 1D tungsten rod is 1.2 μm wide, the rod-to-rod spacing is 4.2 μm and the filling fraction of tungsten material is 28%. The bottom 'V' groove is formed owing to a slow KOH etching $\{111\}$ planes of the $\langle 001 \rangle$ oriented silicon substrate. The step coverage of the deposition process is not 100%, and this gives rise to the formation of a keyhole in the centre of the more deeply imbedded lines.

oscillations in the photonic allowed band.

To confirm our experimental observation, a transfer-matrix calculation⁴ of transmittance and reflectance was carried out, and the results are shown in Fig. 3. The structure parameters used are the same as the fabricated structure except that we used a slightly higher filling fraction, 33.3%. Frequency-dependent dielectric functions $\epsilon_1(\omega)$, $\epsilon_2(\omega)$ for tungsten were also used, to take into account dispersion and absorption effects^{16,17}.

Figure 3a shows the computed result for a four-layer ($N = 4$) 3D crystal, which correctly predicts the photonic band edge position and the gap size. More importantly, the transmittance (blue curve) also shows high transmission at $\lambda \approx 5 \mu\text{m}$ —although the computed peak value (80%) is about three times higher, and the full-width at half-maximum ($\sim 0.5 \mu\text{m}$) three times narrower, than the measured values. By increasing N to 6 (that is, a sample thickness of $9.6 \mu\text{m}$), the peak transmission reaches a stable value of $\sim 95\%$, further confirming the notion of a photonic crystal. We note that there exists a sharp absorption peak (red curve) at $\lambda \approx 6\text{--}7 \mu\text{m}$. The peak wavelength is near the band edge, and its absorptance is $\sim 50\%$. Using a photoacoustic spectroscopy technique¹⁸, a direct absorption measurement was performed and a clear absorption peak of

22% was observed at $\lambda \approx 6 \mu\text{m}$. For comparison, the absorptance of a uniform tungsten film at this wavelength was measured to be $\sim 1\%$. For $\lambda < 4.5 \mu\text{m}$, computed reflectance ($>75\%$) is much higher than that observed (20–30%). As metallic absorption loss is not important in this wavelength range¹⁶, the observed low R is attributed to scattering losses. One source of scattering is structure imperfections, such as the ‘keyholes’ shown in Fig. 1.

The photonic bandgap attenuation in a 3D metallic photonic crystal must not be confused with typical metallic attenuation. To illustrate this point, transmission spectra for 3D crystals of different N values (2, 4 and 6) were computed, and the results are shown in Fig. 3b. The dashed line is a reference spectrum taken from a uniform 6,000-Å tungsten film. Consistent with its small metallic skin depth (300–500 Å for $1 \mu\text{m} < \lambda < 25 \mu\text{m}$), the tungsten film transmittance is very low ($T < 10^{-8}$) and is nearly independent of wavelength. The sample spectrum, on the other hand, exhibits a much higher transmission ($T \approx 10^{-1}$) for $\lambda < 6 \mu\text{m}$, suggesting that photonic transport in this spectral range is not dominated by metallic attenuation. Moreover, a strong wavelength-independence and N -dependence is observed in the bandgap regime ($\lambda > 8 \mu\text{m}$). This N -dependence indicates that transmittance attenuation at $\lambda > 8 \mu\text{m}$ scales with the layer thickness of our 3D structure, but not with metallic skin depth. Thus, the attenuation at $\lambda > 8 \mu\text{m}$ is due primarily to the photonic bandgap effect. The attenuation

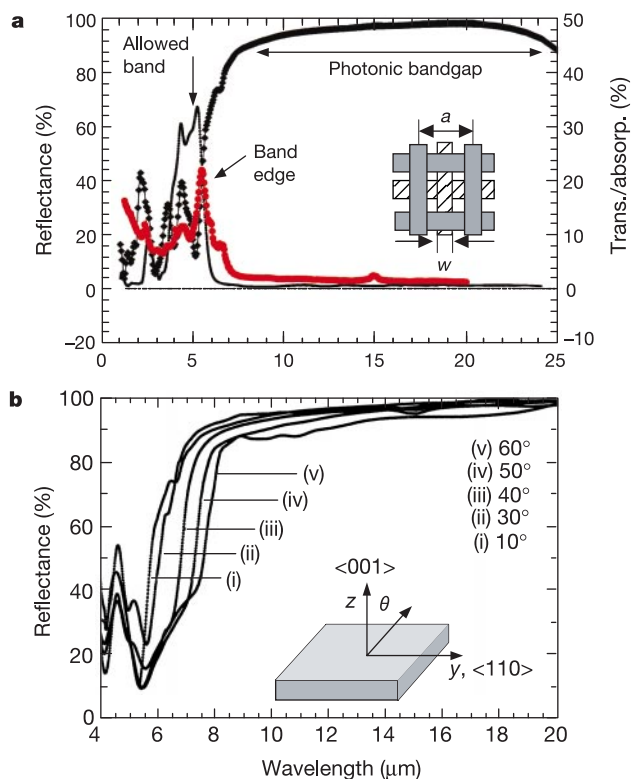


Figure 2 Measured reflectance and transmission/absorption spectra of the tungsten photonic crystal. **a**, The measured reflectance (black diamonds), transmission (blue circles) and absorptance (red circles) spectra for light propagating along the $\langle 001 \rangle$ axis. The band edge occurs at $\lambda \approx 5\text{--}6 \mu\text{m}$. The simultaneous observation of a high reflectance and a low transmittance for $\lambda > 6 \mu\text{m}$ is indicative of a large metallic photonic bandgap. A clear absorptance peak is also observed at $\lambda \approx 6 \mu\text{m}$. The high transmission at $\lambda \approx 5 \mu\text{m}$ is attributed to allowed band propagation of light through this otherwise impenetrable metallic mesh. The inset shows a schematic top view of the 3D crystal. **b**, Tilt-angle reflectance spectra taken from the four-layer tungsten photonic crystal. The crystal orientation is tilted from the $\langle 001 \rangle$ to the $\langle 110 \rangle$ axes and the light incident angle (θ) is therefore systematically tilted away from $\Gamma\text{--}X'$ toward $\Gamma\text{--}L$ of the first Brillouin zone. The tilt angles are 10° , 30° , 40° , 50° and 60° . As θ is increased, the band edge moves from $\lambda \approx 6 \mu\text{m}$ at $\theta = 10^\circ$ to $\lambda \approx 8 \mu\text{m}$ at $\theta = 60^\circ$. Despite the shift in band edge, a large absolute photonic bandgap exists, from $\lambda \approx 8 \mu\text{m}$ to $\lambda > 20 \mu\text{m}$, for our 3D tungsten photonic crystal.

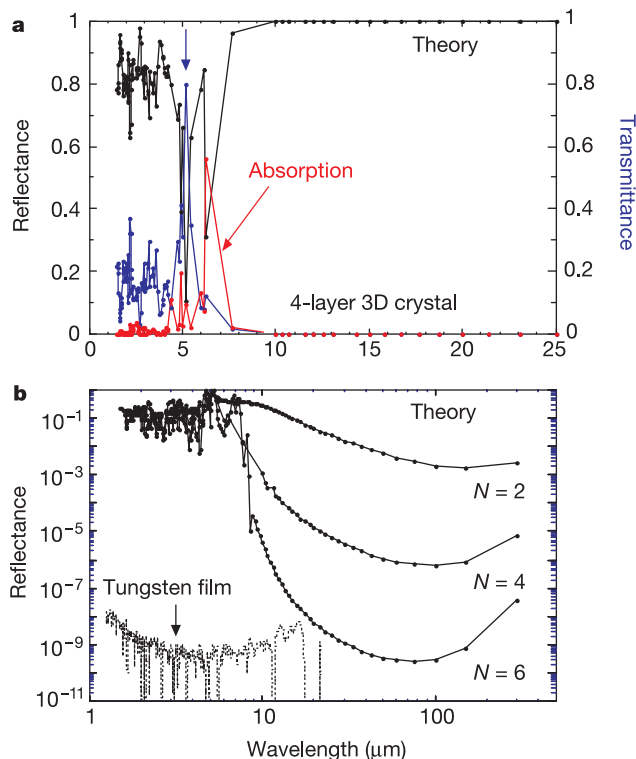


Figure 3 Calculated reflectance and transmission/absorption spectra of 3D tungsten photonic crystals. **a**, Calculated reflectance (black dots), transmittance (blue dots) and absorptance (red dots) spectra for the four-layer 3D tungsten photonic crystal. It correctly predicts the band edge position, the gap size and also the high transmission at $\lambda \approx 5 \mu\text{m}$. A strong absorption peak (50%) is also present at $\lambda \approx 6\text{--}7 \mu\text{m}$, the photonic band edge. **b**, Calculated transmission spectra for 3D tungsten photonic crystals with different numbers of layers: $N = 2, 4$ and 6 . Note the log–log scale. The dashed line is a reference spectrum measured for a uniform 6,000-Å tungsten film. The reference is nearly wavelength-independent, and its amplitude is determined by metallic skin depth. Yet, the sample spectrum has a much higher transmission for $\lambda < 5.5 \mu\text{m}$, and exhibits a strong wavelength-independence and N -dependence for $\lambda > 8 \mu\text{m}$ (that is, in the bandgap). Such an N -dependence suggests further that transmittance attenuation at $\lambda > 8 \mu\text{m}$ is dominated by the photonic bandgap effect.

constant in the photonic bandgap is very large: it is ~ 32 , 56 and 64 dB per unit cell at $\lambda = 10$, 20 and 40 μm , respectively. This means that only one unit cell of a 3D tungsten crystal is sufficient to achieve strong attenuation of electromagnetic waves.

To prove the existence of an absolute metallic photonic bandgap, that is, a common bandgap for light propagating in all directions, a tilt-angle reflection experiment was conducted. Five tilt-angle spectra are shown in Fig. 2b for $\theta = 10^\circ$, 30° , 40° , 50° and 60° , respectively. As θ is increased, the band edge position moves from $\lambda \approx 6 \mu\text{m}$ for $\theta = 10^\circ$ to $\lambda \approx 8 \mu\text{m}$ for $\theta = 60^\circ$. Nonetheless, both the oscillating features at $\lambda < 6 \mu\text{m}$ and the high reflectance at longer wavelengths remain for all values of θ measured. Despite the shift in band edge, a large complete photonic bandgap exists, from $\lambda \approx 8 \mu\text{m}$ to $\lambda > 20 \mu\text{m}$, for our 3D tungsten photonic crystal. Such an unusually large bandgap could be used for suppressing broadband blackbody radiation in the infrared¹⁹, and recycling its energy into the visible spectrum. In the photon recycling process, an absolute 3D photonic bandgap completely frustrates infrared thermal emission, and forces the radiation into a selective emission band. Consequently, energy is not wasted in heat generation, but rather rechannelled into a useful emission band. According to Kirchhoff's law, the integrated absorptance equals integrated emissivity²⁰. The absorption peak near the band edge (see Fig. 3) can then act as a channel for light emission. This energy-recycling

mechanism is based on an interplay between photonic bandgap rejection and photonic band edge enhanced absorption in the 3D metallic crystal. A potential application of this effect is as an incandescent lamp; the tungsten 3D crystal could be heated up to $> 1,500^\circ\text{C}$, and the emission band shifted into the visible region. Another possibility is thermal-photovoltaic (TPV) applications²¹. Using our 3D structure as an emitter, a TPV model calculation²¹ shows that the TPV conversion efficiency reaches 51%, which is to be compared to the 12.6% efficiency for a blackbody emitter. In practice, the photon recycling efficiency may be limited by the broad background observed experimentally in the absorption spectrum (Fig. 2a).

A 3D metallic crystal not only exhibits a strong photonic bandgap, but also possesses unique transmission and absorption characteristics. We note that whereas transmittance through a uniform 6,000-Å tungsten film is extremely small, $T < 10^{-8}$, peak transmission through the 3D tungsten crystal sample at $\lambda \approx 5 \mu\text{m}$ is high, $T \approx 25\text{--}30\%$. We may also approximate our 3D structure as arrays of subwavelength holes, and the estimated transmission efficiency is also small, 3×10^{-5} (ref. 22). Here, the hole radius is estimated from the straight opening shown in Fig. 1a inset. Neither of the above two approximations can explain the observed transmission enhancement. A further transfer-matrix calculation reveals that the transmission peak wavelength scales linearly with lattice constant a , and depends on material filling fraction. A similar enhancement effect and scaling behaviour have also been observed in two-dimensional metallic thin-film hole arrays, and attributed to surface plasma excitation²³. However, its peak wavelength is independent of the hole diameter, or equivalently, hole filling fraction. The surface plasmon picture is useful in describing electromagnetic (EM) modes in thin films with small holes, where the film thickness is small compared with wavelength (ref. 23). It is likely that the plasmon effect will play a role. However, a formal theoretical classification of the nature of the EM modes in complex structures such as ours would give better insights into this phenomenon. Such EM modes must also manifest the 3D crystal symmetry, and facilitate waveguiding through metallic openings²⁴. To explore the manner by which the light wave manages to mould itself according to the intricate 3D metallic structure (and to the Bloch theorem), a finite-difference time-domain (FDTD) calculation was carried out.

FDTD calculations were done for a six-layer 3D crystal sample at three different wavelengths λ (5.2, 6.5 and 12 μm) and at two separate time steps ($T = 1\Delta t = 0.133 \times 10^{-12}$ s and $2\Delta t$). This calculation assumes a perfect metallic boundary condition, as tungsten material absorption is minimal in this wavelength range¹⁶. Here the values of Δt are chosen to display the transient and steady-state electromagnetic wave distribution, respectively. The colour plot (Fig. 4) is expressed on a logarithmic scale, and shows electric intensity profile in the y - z plane. The light source is a continuous-wave point source, indicated by arrows, and is placed at mid-point of the first layer ($L = 1$). The tungsten rods show up as red regions. At $\lambda = 5.2 \mu\text{m}$ and $T = 1\Delta t$ (Fig. 4a), light-wave intensity (the dark blue colour) diverges in the y direction, and meanwhile its wavefront propagates along the z axis up to $L = 6$. At $T = 2\Delta t$ (Fig. 4b), light has travelled uniformly through all six layers and its intensity has reached a steady state. At the metallic boundaries, there exists a strong field gradient—from red to green to blue—encompassing the metallic rods. These periodic metallic–air boundaries mould the flow of light and dictate its Bloch-wave transport characteristics. At $\lambda = 6.5 \mu\text{m}$ (Fig. 4c and d), light also diverges in the y direction, propagates less intensively toward $L = 4$ at $T = 1\Delta t$ and eventually through $L = 6$ at $T = 2\Delta t$. Additionally, the speed of light slows down considerably at this wavelength. The weaker transmission and the retarded speed of light are clear manifestations of light propagation at a photonic band edge. This retardation, along with percolation of light through the 3D structure, may be responsible

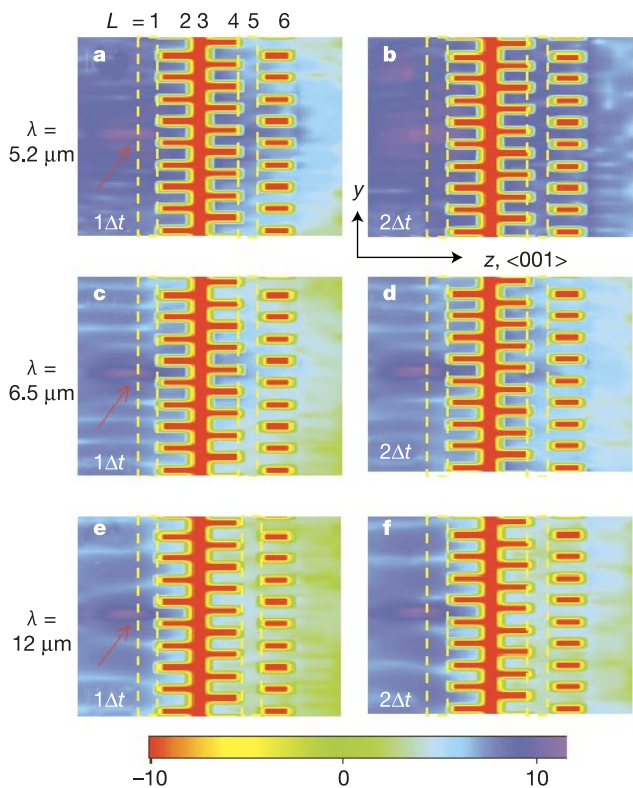


Figure 4 Results of a finite-difference time-domain (FDTD) calculation for a six-layer tungsten 3D photonic crystal. Calculations were done for three different wavelengths ($\lambda = 5.2$, 6.5 and 12 μm) and at two separate time steps ($T = 1\Delta t$ and $2\Delta t$). The simulation light source is a point source, indicated by arrows, and intensity profile is plotted across the y - z plane on a logarithmic scale. The purple and red colours represent the strongest and weakest intensity, respectively. The y - z plane intersects with even-layer metallic rods ($L = 2, 4$ and 6) and the $L = 3$ odd-layer rod. The two yellow dashed lines indicate locations of the other two odd-layer rods, $L = 1$ and 5 . The light wave travels strongly through the 3D crystal at $\lambda = 5.2 \mu\text{m}$ (a, b), less intensively and more slowly at $\lambda = 6.5 \mu\text{m}$ (c, d) and is totally reflected at $\lambda = 12 \mu\text{m}$ (e, f), consistent with the behaviour of light propagation in the photonic allowed-band, band-edge and forbidden bandgap.

for the large absorption at $\lambda = 6\text{--}7\text{ }\mu\text{m}$. Here the unique photonic band edge behaviour leads to absorption enhancement by orders of magnitude. At $\lambda = 12\text{ }\mu\text{m}$ (Fig. 4e and f), light does not travel through the six-layer structure and is totally reflected back, as it is in the photonic bandgap region. □

Methods

Creation of a 3D single-crystal metallic photonic crystal

Details of the fabrication of the polysilicon/SiO₂ 3D photonic crystal, which formed the basis of the tungsten moulds, are given in ref. 15. Briefly, the 3D silicon photonic crystal consists of layers of one-dimensional rods with a stacking sequence that repeats itself every four layers (a unit cell), and has a face-centre-tetragonal lattice symmetry²⁵. Further details of the rest of the processing steps are as follows. First, the polysilicon in a 3D silicon photonic crystal was removed using a 6 M, 85 °C KOH etch which has a selectivity of ~100:1. Over-etch during the KOH process, which is required to ensure the removal of all the polysilicon, results in the formation of a 'V' structure on the bottom of the layer contacting the substrate. This is due to etching of the underlying substrate. The KOH etch effectively stops when the etch-front encounters the slow-etching 111 planes of the substrate, thus forming a 'V' groove (Fig. 1). But this artefact does not appear to significantly affect photonic bandgap performance. Second, the blanket CVD tungsten film does not adhere to silicon dioxide, and was therefore grown on a 50-nm-thick TiN adhesion layer deposited by reactive ion sputtering. The bulk of the tungsten film was deposited at high pressure (90 torr) from WF₆ and H₂. The CVD of tungsten results in films of very high purity; the film resistivity was 10 $\mu\Omega\text{ cm}$. The step coverage of the deposition process is not 100%, and this gives rise to the formation of a 'keyhole' in the centre of the more deeply imbedded lines (Fig. 1b). However, the resulting thickness is far greater than the skin depth of tungsten, and the parts typically retain sufficient structural integrity to be handled readily. Last, excess tungsten on the surface was removed by chemical mechanical polishing, and the oxide mould was removed with a 1:1 HF solution, which etches SiO₂ but not tungsten or TiN. All of the techniques used throughout are modifications of standard CMOS processes, and all work was performed on commercially available, monitor-grade, six-inch silicon wafers.

Received 26 September 2001; accepted 26 February 2002.

- Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
- John, S. Electromagnetic absorption in a disordered medium near a photon mobility edge. *Phys. Rev. Lett.* **53**, 2169–2172 (1984).
- Genack, A. & Garcia, N. Observation of photon localization in a three-dimensional periodic array. *Phys. Rev. Lett.* **66**, 2063–2067 (1991).
- Sigalas, M. M., Chan, C. T., Ho, K. M. & Soukoulis, C. M. Metallic photonic band-gap materials. *Phys. Rev. B* **52**, 11744–11751 (1995).
- Fan, S., Villeneuve, P. R. & Joannopoulos, J. D. Large omnidirectional band gaps in metalodielectric photonic crystals. *Phys. Rev. B* **54**, 11245–11251 (1996).
- McIntosh, K. A. *et al.* Three-dimensional metalodielectric photonic crystals exhibiting resonant infrared stop bands. *Appl. Phys. Lett.* **70**, 2937–2939 (1997).
- Sievenpiper, D. F., Sickmiller, M. E. & Yablonovitch, E. 3D wire mesh photonic crystals. *Phys. Rev. Lett.* **76**, 2480–2483 (1996).
- Moroz, A. Three-dimensional complete photonic-band-gap structures in the visible. *Phys. Rev. Lett.* **83**, 5274–5277 (1999).
- Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
- Palik, E. D. (ed.) *Handbook of Optical Constants of Solids* 275–409 (Academic, San Diego, 1998).
- Ozbay, E. *et al.* Defect structures in metallic photonic crystals. *Appl. Phys. Lett.* **69**, 3797–3799 (1996).
- Sievenpiper, D. F. *et al.* 3D metallo-dielectric photonic crystals with strong capacitive coupling between metallic islands. *Phys. Rev. Lett.* **80**, 2829–2832 (1998).
- Velev, O. D. & Kaler, E. W. Structured porous materials via colloidal crystal templating: from inorganic oxides to metals. *Adv. Mater.* **12**, 531–534 (2000).
- Zakhidov, A. A. *et al.* Three-dimensionally periodic conductive nanostructures: network versus cermet topologies for metallic PBG. *Synth. Met.* **116**, 419–426 (2001).
- Lin, S. Y. *et al.* A three-dimensional photonic crystal in the infrared wavelengths. *Nature* **394**, 252–253 (1998).
- Ordal, M. A. *et al.* Optical properties of the metals Al, Co, Cu, Au, Fe, Pb, Ni, Pd, Pt, Ag, Ti, and W in the infrared and far infrared. *Appl. Opt.* **22**, 1099–1119 (1983).
- El-Kady, I., Sigalas, M. M., Biswas, R., Ho, K. M. & Soukoulis, C. M. Metallic photonic crystals at optical wavelengths. *Phys. Rev. B* **62**, 15299–15301 (2000).
- McClelland, J. F., Jones, R. W., Lou, S. & Seaverson, L. M. in *Practical Sampling Techniques for Infrared Analysis* (ed. Coleman, P. B.) Ch. 5 (CRC, Boca Raton, Florida, 1993).
- Lin, S. Y., Fleming, J. G., Chow, E. & Bur, J. Enhancement and suppression of thermal emission by a three-dimensional photonic crystal. *Phys. Rev. B* **62**, R2243–R2246 (2000).
- Dereniak, E. L. & Boreman, G. D. *Infrared Detectors and Systems* 74 (Wiley & Sons, New York, 1996).
- Zenker, M., Heinzel, M., Stollwerck, G., Ferber, J. & Luther, J. Efficiency and power density potential of combustion-driven thermophotovoltaic systems using GaSb photovoltaic cells. *IEEE Trans. Elec. Dev.* **48**, 367–376 (2001).
- Bethe, H. A. Theory of diffraction by small holes. *Phys. Rev.* **66**, 163–182 (1944).
- Ebbesen, T. W., Lezec, H. J., Ghaemi, H. F., Thio, T. & Wolff, P. A. Extraordinary optical transmission through sub-wavelength hole arrays. *Nature* **391**, 667–669 (1998).
- Porto, J. A., Garcia-Vidal, F. J. & Pendry, J. B. Transmission resonance on metallic gratings with very narrow slits. *Phys. Rev. Lett.* **83**, 2845–2848 (1999).
- Ho, K. M. *et al.* Photonic band gap in three-dimensions: new layer-by-layer periodic structure. *Solid State Commun.* **89**, 413–416 (1994).

Acknowledgements

We thank J. Gees and J. Moreno for discussions, and M. Tuck and J. Bur for technical support. The work at Sandia National Laboratories was supported by the US DOE. Sandia is a multi-programme laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US DOE.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.Y.L. (e-mail: slin@sandia.gov).

Single-substrate liquid-crystal displays by photo-enforced stratification

Roel Penterman*, Stephen I. Klink*, Henk de Koning*, Giovanni Nisato* & Dirk J. Broer*†

* Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

† Department of Polymer Chemistry and Technology, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands

Data visualization plays a crucial role in our society, as illustrated by the many displays that surround us. In the future, displays may become even more pervasive, ranging from individually addressable image-rendering wall hangings to data displays integrated in clothes¹. Liquid-crystal displays (LCDs) provide most of the flat-panel displays currently used. To keep pace with the ever-increasing possibilities afforded by developments in information technology, we need to develop manufacturing

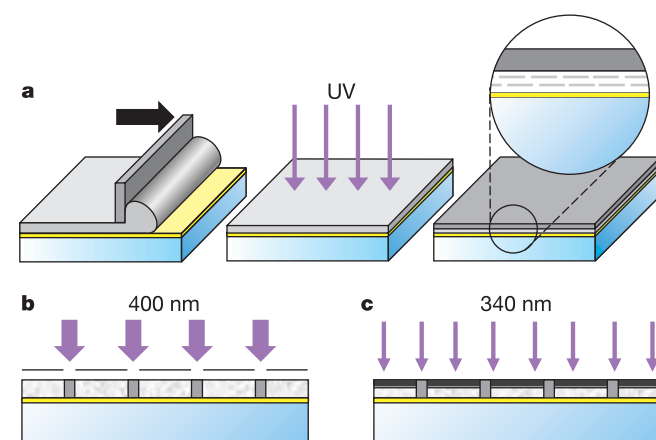


Figure 1 Schematic representation of the stratification process. The basic principle consists of the formation of two layers from a 20- μm film consisting of a blend of isobornylmethacrylate, stilbene-dimethacrylate and E7 by photopolymerization-induced phase separation in the vertical direction (a). In order to couple the polymer top film mechanically to the bottom substrate, phase separation is established by a two-step ultraviolet exposure. During the first exposure (b), to ultraviolet radiation of 400 nm wavelength, 100- μm -wide walls are polymerized, forming the sides of individual boxes still containing a mixture of LC and pre-polymer. During the second exposure (c), to ultraviolet radiation of 340 nm wavelength, the 10- μm -thick cover sealing the boxes is polymerized.

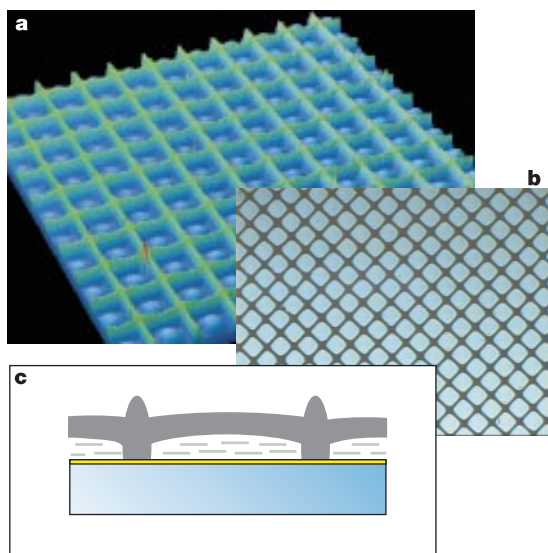


Figure 2 Shapes of the liquid-crystal-filled boxes of the PES-LCD, as observed by interferometry (**a**) and through a polarizing microscope (**b**). Interferometry of the surface reveals elevations of a few micrometres at the position of the walls. A schematic cross-section of the individual stratified boxes ($500\ \mu\text{m} \times 500\ \mu\text{m} \times 20\ \mu\text{m}$) is shown in **c**. The thickness of the LC is about $10\ \mu\text{m}$. Note that in **a** and **c** the vertical dimensions are much expanded with respect to the horizontal dimensions.

processes that will make LCDs cheaper and larger, with more freedom in design. Existing batch processes for making and filling LCD cells^{2,3} are relatively expensive, with size and shape limitations. Here we report a cost-effective, single-substrate technique in which a coated film is transformed into a polymer-covered liquid-crystal layer. This approach is based on photo-enforced stratification: a two-step photopolymerization-induced phase separation of a liquid-crystal blend and a polymer precursor. The process leads to the formation of micrometre-sized containers filled with a switchable liquid-crystal phase. In this way, displays can be produced on a variety of substrates using current coating technology. The developed process may be an important step towards new technologies such as 'display-on anything' and 'paintable displays'.

The basis for the photo-enforced stratification (PES) LCD technology is a directionally controlled photopolymerization-induced phase separation of a homogeneous mixture of a liquid crystal and a polymer-forming material. Polymerization-induced phase separation has been utilized to produce polymer dispersed liquid-crystal displays^{4–6} and other electro-optical devices of complex geometries^{7–10}. These processes are normally carried out in glass or plastic cells, where interfacial properties may enhance the formation of layers at the appropriate surfaces^{9,10}. Directionality of the phase separation as a result of photoinduced diffusion can be achieved by using the absorption of a compound in the mixture to provide a gradient of the actinic radiation over the layer thickness^{10–12}. In this study we have realized the formation of phase-separated layers by the photopolymerization of coated films on a single substrate in a nitrogen atmosphere. The mixture of liquid crystal and monomer is applied as a thin film onto a substrate that is provided with interdigitated electrodes and a polyimide orientation layer. The typical film thickness is in the range of tens of micrometres; known film-forming techniques for fast coating of both small and large areas, such as the doctor blade coating technique or slot die coating, may be used to apply the film. Upon exposure to ultraviolet radiation with a selected wavelength, the film very accurately separates into the pure liquid-crystal layer at the bottom and a hard polymer coating at the top. This step is referred to as

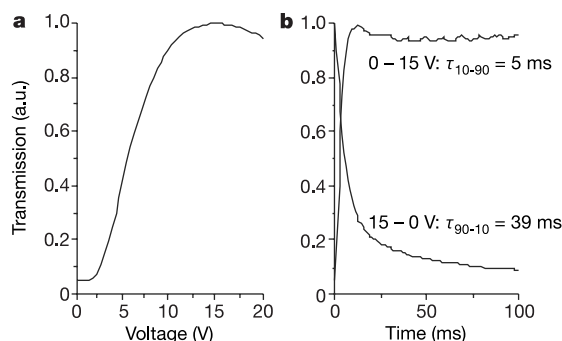


Figure 3 Switching behaviour of a PES-LCD. Transmission voltage (**a**) and switching rate (**b**) measured in transmission. The liquid crystal (E7) is driven by in-plane switching. The indium tin oxide electrodes are separated with a spacing of $9\ \mu\text{m}$, while the electrode pitch is $18\ \mu\text{m}$. (a.u., arbitrary units.) τ_{10-90} is the time that the sample switches from 10 to 90% relative transmission; τ_{90-10} is the reversed switching time.

stratification.

The processing sequence is illustrated in Fig. 1a. The liquid crystal typically consists of a blend of low-molar-mass rod-shaped molecules. For this typical experiment E7 (Merck) has been selected, which is a blend of a cyanoterphenyl and several cyanobiphenyl molecules. The monomer mixture has been chosen such that in its monomeric state it dissolves in the liquid crystal, but phase-separates in a stratified manner during polymerization. Isobornyl-methacrylate is the preferred monomer, as it combines properties such as high solubility and low viscosity before polymerization, high diffusion rates and controlled phase separation upon polymerization, and rigidity after polymerization. In order to have a better control over the spatial position of the phase-separated polymer layer, a stilbene-functionalized dimethacrylate monomer has been added to provide an intensity gradient over the film thickness for ultraviolet radiation in the wavelength region around $340\ \text{nm}$, which is used to irradiate the film during the stratification step. As a result, polymerization predominantly takes place near the liquid/nitrogen interface that is directed towards the ultraviolet source. Moreover, the stilbene-dimethacrylate monomers provide crosslinks between the poly(isobornylmethacrylate) chains, which further enhances the directionally controlled phase-separation process. Table 1 shows a typical composition of the stratifying layer.

In order to obtain a stable device that can withstand lateral forces, it is essential that the polymer topcoat and the substrate be mechanically coupled. This is realized by using a two-step ultraviolet exposure for the PES process. The blend of monomer and liquid crystal is first exposed through a mask with high-intensity ultraviolet radiation of $400\ \text{nm}$ wavelength, which is outside the absorption region of the stilbene-dimethacrylate. Polymerization in the exposed areas leads to horizontal stratification, and this results in the formation of polymer walls (Fig. 1b). Subsequently, the previously unexposed areas are cured with low-intensity ultraviolet radiation of $340\ \text{nm}$ wavelength, at which the stilbene-dimethacrylate exhibits significant absorption. This induces the stratification process in the vertical direction (Fig. 1c). The whole process finally leads to the formation of polymer boxes filled with pure liquid crystal. The boxes consist of walls formed during the horizontal stratification step, and the lid of each box is formed during the vertical stratification step. In the examples shown here, the boxes are $500 \times 500\ \mu\text{m}$ wide and $10\ \mu\text{m}$ high, the walls are about $100\ \mu\text{m}$ wide and the lids are $10\ \mu\text{m}$ thick.

The resulting structure has been visualized by interferometric microscopy of the surface of the top layer (Fig. 2a). The liquid crystal was chosen to follow a planar director profile. Figure 2b shows a polarization microscope picture of a PES-LCD placed at 45°

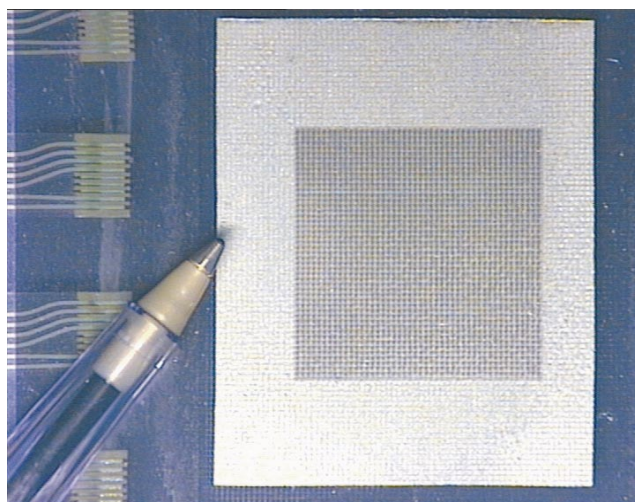


Figure 4 Photograph of a PES-LCD demonstration device. This 24-pixel device is operated in the reflective mode by using a polarizing reflector at its back. Both back and front polarizer are placed parallel to the in-plane switching electrodes to obtain a normally

white mode. The whole active pixel area of $2.5 \times 2.5 \text{ cm}^2$ is addressed, and appears dark. The end of a ballpoint pen is shown for scale.

with respect to the crossed polarizers. By analysing Fig. 2a and b, a schematic cross-section (Fig. 2c) can be deduced. At the base of each box, a set of previously structured interdigitated electrodes switch the liquid crystal by in-plane electrical fields^{13,14}. The selected in-plane switching liquid-crystal effect combines fast switching (ranging from 5 to 40 ms) and sufficient contrast (of the order of 1:20). Typical switching characteristics are shown in Fig. 3.

A photograph of a display produced by the twofold stratification process is shown in Fig. 4. Here the PES-LCD is provided with a top polarizer, and has a diffusive reflector/polarizer combination at the back in order to use the display in a reflective mode. The whole display thickness is determined by the sum of the thickness of a single substrate, which can be as thin as 0.5 mm, and the thickness of the stratified layer, which is less than $30 \mu\text{m}$. When applied on a plastic substrate, the display can be even thinner ($<150 \mu\text{m}$) and can be made flexible. In line with the concept of paintable displays, the PES-LCD can be coated with two 600-nm polarizer films that are applied by shear-orientation of a sulphonated perylene dye, doctor

bladed from a water solution¹⁵.

Besides the many advantages that can be ascribed to the PES display technology, a number of limitations can also be recognized. There will be some aperture reduction, leading to a lower front-of-screen brightness, because of the presence of the polymer walls that provide the mechanical stability. These walls therefore need to be placed at the inactive areas of the displays—for example, at pixel edges. In addition, the contrast of the display remains limited, because the thickness and the shape of the liquid-crystal areas have not been optimized yet. Preliminary experiments have shown that much flatter polymer/liquid-crystal interfaces can be obtained by improved exposure conditions (collimation, intensity distribution) and materials (nonlinear response to ultraviolet radiation by added inhibitor). Another issue that might affect the acceptance of this new technology is its sensitivity towards mechanical forces exerted on the top layer. In order to improve the robustness of these PES displays, they could be provided with a hard top coating—for instance, on top of the coated polarizer layer. □

Table 1 Composition of the stratifying layer

Component	Chemical structure	wt%
E7 liquid crystal blend		50.0
Isobornylmethacrylate		44.5
Stilbene-dimethacrylate		5.0
Photoinitiator Irgacure 651		0.5

Received 7 January; accepted 4 March 2002.

1. Farrington, J., Moore, A. J., Tilbury, N., Church, J. & Biemond, P. D. in *Proc. IEEE 3rd Int. Symp. on Wearable Computers 1999* 107–113 (IEEE, San Francisco, 1999).
2. Morozumi, S. in *Liquid Crystals: Applications and Uses* (ed. Bahadur, B.) Vol. 1, 181 (World Scientific, Singapore, 1990).
3. Kamiya, H. *et al.* in *SID 01 Digest of Technical Papers, 2001 Int. Symp.* Vol. 32, 1354–1357 (Society for Information Display, 2001).
4. Vaz, N. A., Smith, G. W. & Montgomery, G. P. A light control film composed of liquid crystal droplets in a UV-curable polymer. *Mol. Cryst. Liq. Cryst.* **146**, 1–15 (1987).
5. Doane, J. W., Vaz, N. S., Wu, B. G. & Zumer, S. Field controlled light scattering from nematic microdroplets. *Appl. Phys. Lett.* **48**, 269–271 (1986).
6. Hirai, Y., Niyama, S., Kumaim, H. & Gunjima, T. Phase diagram and phase separation in LC/prepolymer mixture. *Proc. SPIE Int. Soc. Opt. Eng.* **1257**, 2–8 (1990).
7. Bowley, C. C., Yuan, H. & Crawford, G. P. Morphology of holographically-formed polymer dispersed liquid crystals (H-PDLC). *Mol. Cryst. Liq. Cryst. Technol. A* **331**, 2069–2076 (1999).
8. Yamada, N., Kohzaki, S., Funada, F. & Awane, K. Axially symmetric aligned microcell (ASM) mode: electro-optical characteristics or new display mode with excellent wide viewing angle. *J. Soc. Inform. Disp.* **3**, 155–158 (1995).
9. Park, E. Y., Taheri, B., West, J. L. & Palfy-Muhoray, P. in *SID 00 Digest of Technical Papers, 2000 Int. Symp.* Vol. 31, 782–785 (Society for Information Display, 2000).
10. Vorflusev, V. & Kumar, S. Phase-separated composite films for liquid crystal displays. *Science* **283**, 1903–1905 (1999).
11. Qian, T., Kim, J.-H., Kumar, S. & Taylor, P. L. Phase-separated composite films: Experiment and theory. *Phys. Rev. E* **61**, 4007–4010 (2000).
12. Broer, D. J., Lub, J. & Mol, G. N. Wide-band reflective polarizers from cholesteric polymer networks with a pitch gradient. *Nature* **378**, 467–469 (1995).
13. Kiefer, R., Weber, B., Windscheid, F. & Baur, G. in *Proc. 12th Int. Display Conf., Japan Display '92* 547–550 (1992).
14. Oh-e, M., Ohta, M., Aratani, S. & Kondo, K. in *Proc. 15th Int. Display Conf. Asia Display '95* 577–580 (1995).
15. Bobrov, Y. A. *et al.* Novel dichroic polarizing materials and approaches to large-area processing. *Mater. Res. Soc. Symp. Proc.* **508**, 225–228 (1998).

Acknowledgements

We thank J. Lub for synthesis of the stilbene dimethacrylate and for discussions.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.B. (e-mail: dick.broer@philips.com).

Distribution of breaking waves at the ocean surface

W. Kendall Melville & Peter Matusov

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0213, USA

Surface waves play an important role in the exchange of mass, momentum and energy between the atmosphere and the ocean. The development of the wave field depends on wind, wave-wave and wave-current interactions and wave dissipation owing to breaking, which is accompanied by momentum fluxes from waves to currents. Wave breaking supports air-sea fluxes of heat and gas^{1,2}, which have a profound effect on weather and climate. But wave breaking is poorly quantified and understood. Here we present measurements of wave breaking, using aerial imaging and analysis, and provide a statistical description of related sea-surface processes. We find that the distribution of the length of breaking fronts per unit area of sea surface is proportional to the cube of the wind speed and that, within the measured range of the speed of the wave fronts, the length of breaking fronts per unit area is an exponential function of the speed of the front. We also find that the fraction of the ocean surface mixed by breaking waves, which is important for air-sea

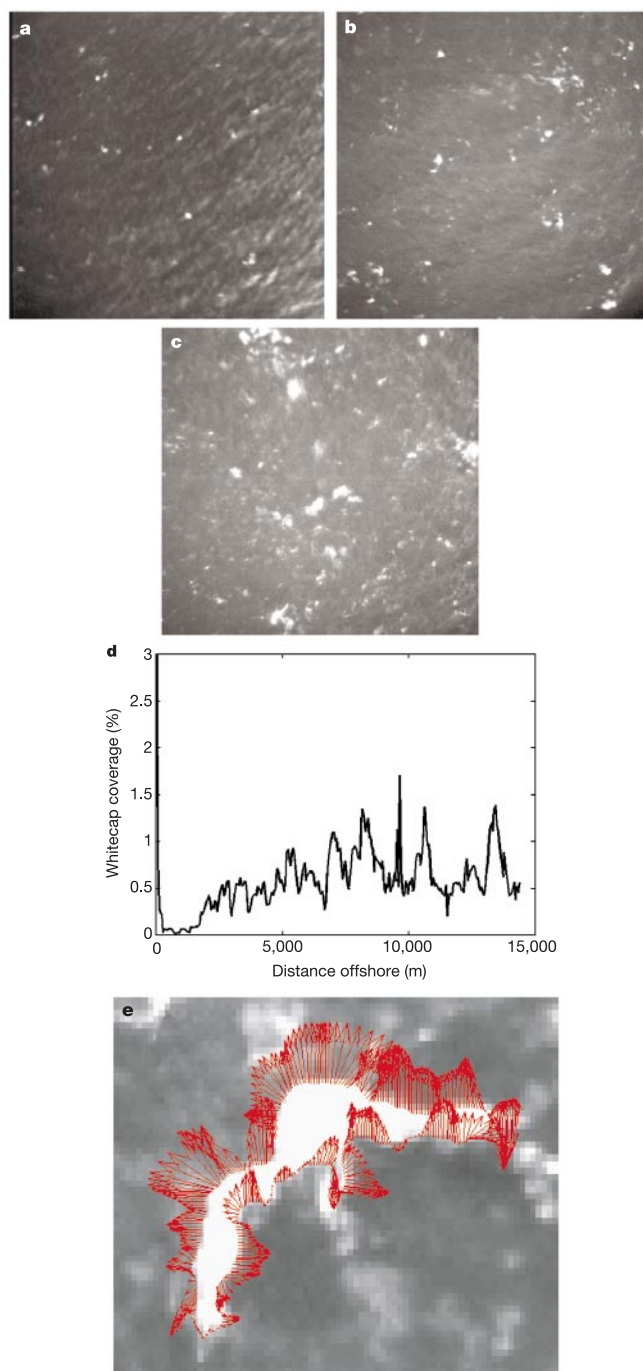


Figure 1 Whitecap imaging and image processing. **a–c**, Cropped images showing approximately $160\text{ m} \times 160\text{ m}$ of the sea surface taken from an altitude of approximately 400 m for mean wind speeds, U_{10} , of 7.2, 9.8 and 13.6 m s^{-1} , respectively, off the coast of North Carolina in the autumn of 1999. We note the increasing density of whitecaps with wind speed and their random shapes. Individual whitecaps are defined by a brightness threshold, which was varied to ensure that the results presented are not sensitive to the threshold chosen. **d**, The fractional area of the sea surface covered by whitecaps, A_w , ('whitecap coverage') as the aircraft flew offshore for a distance of approximately 15 km. We note the initial decline in A_w across the surf zone and then the increase offshore. There are large fluctuations at spatial scales of $O(1\text{--}10\text{ km})$ that may be due to both local wind and wave modulations. **e**, The result of processing a single whitecap using the PIV technique in which the normal velocity of the boundary of the whitecap is resolved at a scale of approximately 50 cm. The mean velocity of the whitecap was computed from the mean of the elemental velocities, and then only the elements located at the boundary within $\pm 90^\circ$ from the mean velocity of the whitecap, having positive outward normal velocities directed outside the boundary, were included in the $\Delta(c)$ statistics.

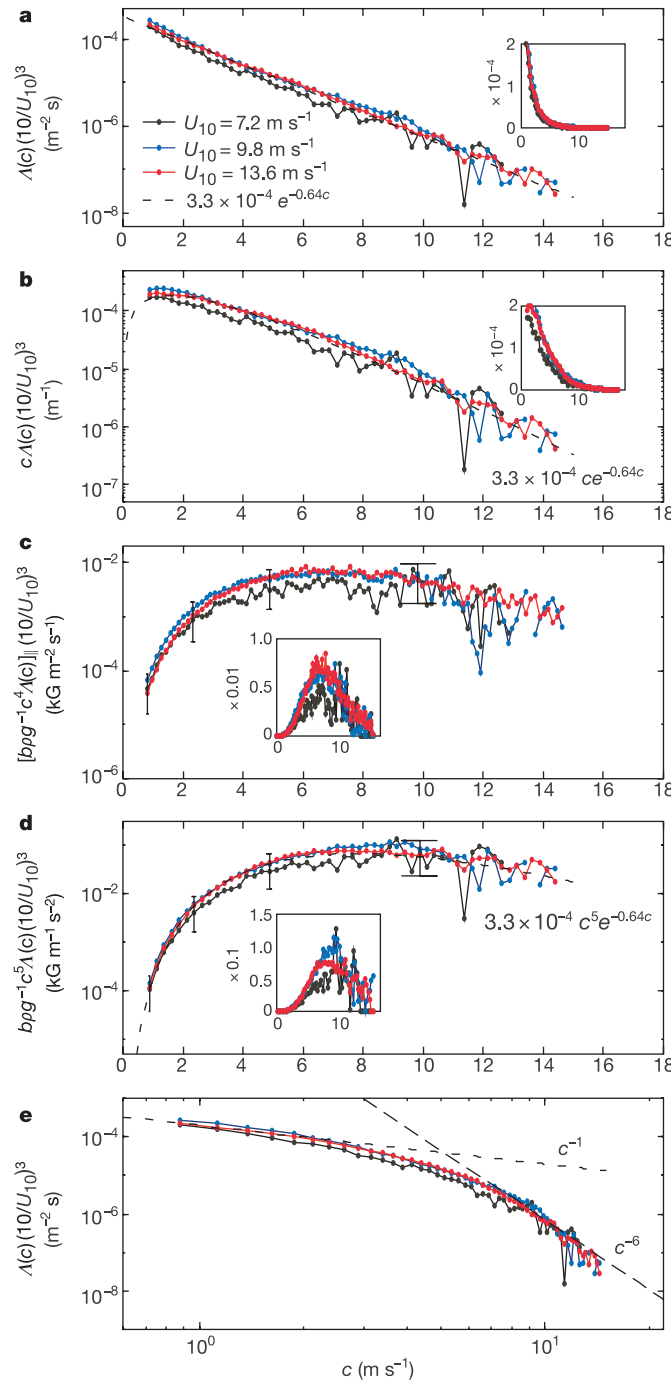


Figure 2 Measurements of $\Delta(c)$ and its moments. **a**, Binned measurements of $\Delta(c)$ (from data like that in Fig. 1e, weighted by U_{10}^{-3}) showing that to a very good approximation the average length of breaking fronts in $(c, c + dc)$ per unit area of sea surface increases like U_{10}^3 and decays exponentially with c as follows: $\Delta(c) = 3.3 \times 10^{-4} e^{-0.64c}$. This PIV method was checked against simpler measurements of the rate of increase of area of individual whitecaps. The line integral of the outward flux vectors around the perimeter of the whitecap is equal to the rate of increase of area of the whitecap, which can easily be determined by other methods. It was found that if the length of the breaking front was on average set to one-third of the perimeter, then $\Delta(c)$ determined by the two methods agreed very well, especially for the smaller values of c . Linear plots of the data (insets) show that the U_{10}^3 scaling is not an artefact of the log-linear plot. See also **b** and **c**. **b**, The weighted (by U_{10}^{-3}) first moment of $\Delta(c)$ which corresponds to the fractional area swept out (turned over) by breakers in the speed range $(c, c + dc)$. This and higher moments are computed from the raw data rather than just multiplying the data in **a** by c^n , although the moments based on the exponential fit to the binned values of $\Delta(c)$ agree very well. The first moment approaches a maximum at smaller c , demonstrating the significance of the

slower/smaller scales in breaking/mixing the surface. **c**, The weighted (by U_{10}^{-3}) fourth moment of $\Delta(c)$, $c^4 \Delta(c)_{||}$, which corresponds to the momentum flux from waves to currents due to whitecaps, with the subscript '||' referring to the component in the wind direction. Now the momentum flux is a maximum in the intermediate range of breaking speeds. Vertical bars show the range of b from laboratory data⁵. **d**, The weighted (by U_{10}^{-3}) fifth moment of $\Delta(c)$, which corresponds to the energy lost from the wave field owing to breaking: 'wave-dissipation'. Again, the weighted curves collapse very well at the smaller values of c , and the maximum moves to higher c for the higher moment. In **c** and **d**, we have used the numerical value of b determined in laboratory experiments on unsteady breaking⁵ (8.5×10^{-3}) and the vertical error bars reflect the range of that parameter. The linear insets in **c** and **d** show that the integrals of these moments would converge over the range of the experiments. **e**, The same data as in **a** but now plotted as log-log, showing that $\Delta(c)$ varies like c^{-1} and c^{-6} for the smallest and largest values of c , respectively. The c^{-1} and c^{-6} behaviour are consistent with local approximations to the exponential decay of $\Delta(c)$. The c^{-6} -behaviour is consistent with a prediction of local equilibrium by Phillips³.

exchange, is dominated by wave breaking at low velocities and short wavelengths.

Wave breaking is a process that is vital to many aspects of air–sea interaction and remote sensing of the oceans. It limits the height of surface waves, transfers momentum flux from waves to currents and ‘renews’ the surface (a significant aspect of air–sea heat-flux and gas-flux processes). Wave breaking also generates marine aerosols—which are important in biogeochemistry, cloud physics, atmospheric radiation balances and hurricane thermodynamics—and entrains bubbles, which modify the optical properties of the ocean surface and thereby the interpretation of ocean colour. Our ability to quantify these processes has been inhibited by the absence of good quantitative measures of the distribution of breaking.

Phillips³ introduced $\Lambda(c)dc$, the average length of breaking crests per unit area of ocean surface travelling at velocities in the range $(c, c + dc)$, as a statistical description of breaking, its kinematics and dynamics. The dynamical description is based in part on laboratory measurements, and in part on physical/dimensional arguments^{4–8} that the energy dissipation per unit length of breaking crest is given by:

$$\epsilon_1 = b\rho_w g^{-1} c^5 \quad (1)$$

where ρ_w is the density of water, g gravity, $c = |c|$, and, for unsteady breaking, b is a number in the range $O(10^{-3}–10^{-2})$ that depends on the steepness of the waves⁵. Radar measurements of Λ suggest that b could be an order of magnitude smaller⁹, but the indirect nature of the measurements may account for some of the difference. Wave energy and momentum densities are related by $M = E/c$, with the momentum flux per unit length of crest given by $m_1 = b\rho_w g^{-1} c^4$. Here we use an airborne video system, along with a motion package and differential global positioning system (GPS), to obtain image sequences of breaking waves in an Earth frame.

Measurements were made from light aircraft flights at approximately 50 m s^{-1} airspeed and 450 m altitude off North Carolina, USA, in 1999 during the shoaling waves experiment (SHOWEX); extending up to 60 km offshore over water of 20–50 m depth. The same ground tracks were flown at altitudes of 15–25 m for wind and wave measurements.

Figure 1a–c shows examples of cropped (256×256 pixels; $160 \times 160 \text{ m}$) images of whitecaps at approximately 0.6-m resolution for the three wind speeds. The increasing fractional area of whitecap coverage, A_w , with wind speed^{10,11} is apparent in Fig. 1. The variability of A_w over scales of the order of a kilometre (Fig. 1d) is remarkable, suggesting that the incidence of breaking may ultimately be related to very local conditions. This variability could come from the local wind, or modulation of the surface wave field itself¹².

Individual whitecaps were tracked and measured for 5–10 s. During that time, their areas could be increasing, at a maximum, or decreasing. Active breaking corresponds to expanding whitecaps. Whitecap decay corresponds to a decrease in size as entrained bubbles rise to the surface^{13,14} and burst to form aerosols. Without detailed measurements of the evolution of the whitecap (see below), it was impossible to associate a unique velocity of advance, c , with each whitecap during its growth. However, by using particle imaging velocimetry (PIV)¹⁵, we could measure the velocity of the local boundary of the whitecap, thereby giving $\Lambda(c)dc$ (see Methods). Figure 1e shows the results for a single large whitecap. Thus $\Lambda(c)dc$ is accumulated from the individual elements of the boundary rather than attributing single values of Λ_i and c_i , say, to each whitecap (see Methods). Data were collected at three averaged ten-metre wind speeds, $U_{10} = 7.2, 9.8$ and 13.6 m s^{-1} , in a series of flights.

Figure 2a shows that when weighted by U_{10}^{-3} (ref. 3), the measurements of $\Lambda(c)$ collapse approximately onto a single exponential curve, $\Lambda(c) = 3.3 \times 10^{-4} e^{-0.64c}$, over the measured range of c . The surface wave field, and hence the breaking statistics, depend on fetch, the distance the wind has blown over the ocean surface,

and other factors, so the apparent dependence on only wind speed here is probably due to the fetch being approximately the same for each measured wind event at this location.

For small values of c , $\Lambda(c)$ varies like c^{-1} and for larger c values like c^{-n} , where $n > 1$. (In general, if $\Lambda(c) = ae^{-bc}$, where a and b are constants, it osculates with the curve $y = dc^{-n}$ ($d = \text{constant}$, $n \geq 1$), at $c = n/b$ when $d = ae^{-n(n/b)^a}$.) $\Lambda(c) \propto c^{-6}$ provides a good approximation for larger c values, consistent with Phillips³ equilibrium subrange (see Fig. 2b). A smaller slope at smaller c is

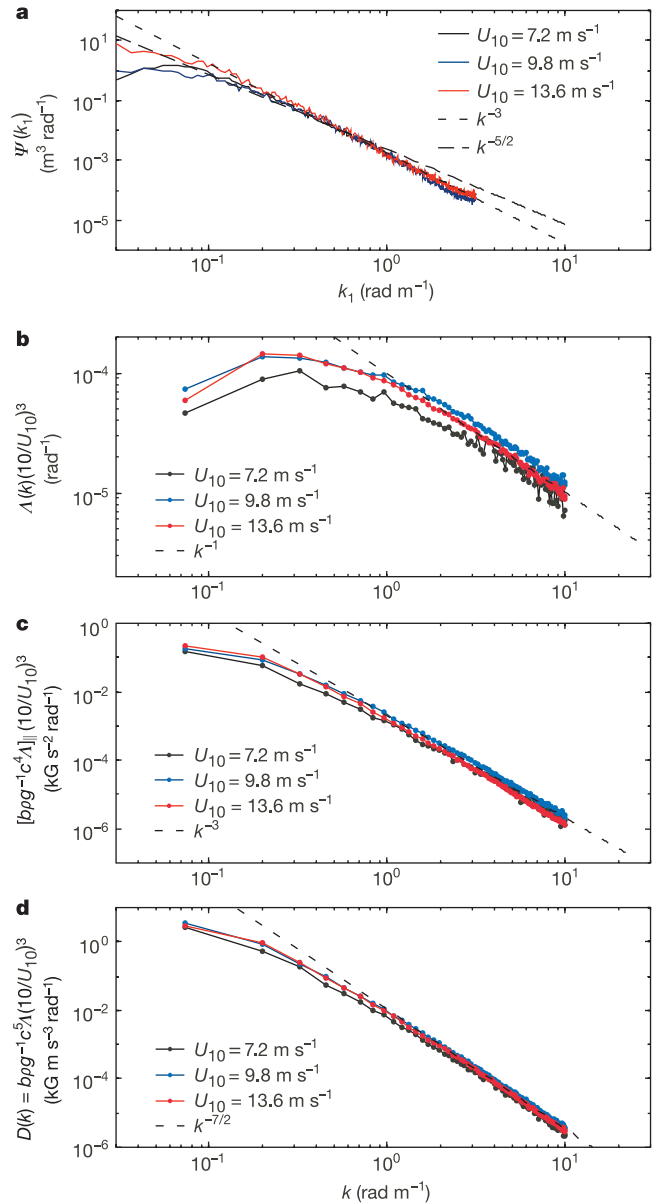


Figure 3 Wavenumber spectra of the surface waves and of $\Delta(c)$ and its moments. **a**, One-dimensional wavenumber spectra of the sea surface displacement measured using laser altimeters on the aircraft in averaged pairs of upwind and downwind flights (one-dimensional samples of the surface) separated briefly in time. Owing to low-wavenumber noise, the data are not considered accurate for wavenumbers below approximately 0.1 rad m^{-1} . Note that wavenumbers $O(0.1) \text{ rad m}^{-1}$ and higher transition from a $k^{-5/2}$ slope to a k^{-3} slope at all wind speeds, consistent with $k^{-5/2}$ and k^{-3} omnidirectional spectra, respectively. **b**, Weighted distribution of $\Delta(k)$ determined from the data of Fig. 2 using $k/\tanh(kh) = g/c_p^2 = (0.8)^2 g/c^2$ to transform from c to k ; $h = 22.5 \pm 2.5 \text{ m}$. For large k the curves decay like k^{-1} . **c**, The downwind component of the weighted momentum flux from breaking waves as a function of wavenumber (compare Fig. 2c). For high wavenumbers the decay is like k^{-3} . **d**, The weighted wave dissipation due to breaking as a function of wavenumber. For high wavenumbers the decay is like $k^{-7/2}$.

qualitatively consistent with the recent equilibrium model of Hara and Belcher¹⁶ (a refinement of Phillips' model), but important quantitative differences remain (see below).

Figure 2c shows the first moment of $\Lambda(c)$ (based on the raw data), where $c\Lambda(c)dc$ is the fraction of the surface area swept out per unit time by breaking waves in the speed range $(c, c + dc)$, and the probability per unit time of breaking at a point on the ocean surface. Given the exponential form for $\Lambda(c)$, we expect a maximum in the n th moment in the neighbourhood of $c = (n/0.64) \text{ m s}^{-1}$. For the first moment this is 1.6 m s^{-1} , close to the lower limit of our measurements.

Using $b = 8.5 \times 10^{-3}$ (ref. 5), Fig. 2d and e shows the weighted (by U_{10}^3) fourth and fifth moments of $\Lambda(c)$, corresponding to the momentum flux from the waves to currents through breaking, and the energy lost by the wave field, respectively. The cubic wind-speed dependence collapses the data for smaller values of c , whereas some statistical scatter due to fewer events remains at larger c . Also shown are insets of the moments, demonstrating that even on a linear scale the U_{10}^3 scaling is robust, especially at the larger wind speeds. Within the scatter at the larger values of c , the integrals of the momentum flux and dissipation distributions converge. The maxima in these moments, inferred from the distribution for $\Lambda(c)$, would be at 6.4 and 8 m s^{-1} , respectively, consistent with the data.

Wave models are usually represented in the wavenumber domain, k . This raises the issue of mapping from c to k . Because breaking is local in space, its spectral representation in the wavenumber plane must be broad-banded. However, measurements suggest that most of the dissipation is around the frequency of the characteristic breaking wave, and that c is approximately 80% of the characteristic

linear phase speed of the wave (see also ref. 9). Thus $k = g/c_p^2 = (0.8)^2 g/c^2$, where $c_p^2 = gk^{-1} \tanh(kh)$ is the phase speed of the waves, h the water depth, and $\Lambda(k)$ and its moments can then be represented, as shown in Fig. 3, along with the one-dimensional wavenumber spectra of the surface waves. Below $k_1 = 0.1 \text{ rad m}^{-1}$ the surface wavenumber spectra are prone to errors, but display power-law behaviour at larger values of k_1 . For $k_1 = (0.1, 1) \text{ rad m}^{-1}$, the spectra are approximately proportional to $k_1^{-5/2}$, consistent with the Phillips³ prediction of a $k^{-7/2}$ two-dimensional spectrum. For $k_1 > 1 \text{ rad m}^{-1}$, the k_1^{-3} one-dimensional wavenumber spectrum corresponds to a two-dimensional k^{-4} spectrum, and a k^{-3} omni-directional spectrum¹⁷. The spectra of $\Lambda(k)$ and the fourth and fifth moments decay like k^{-1} , k^{-3} and $k^{-7/2}$, respectively, for $k > 1 \text{ rad m}^{-1}$.

The directional properties of $\Lambda(c)$, the momentum flux and the energy dissipation are shown for the three different mean wind vectors in Fig. 4. Some elements of the whitecap boundaries are travelling slowly upwind, but when the higher moments are taken, the momentum flux and dissipation are clearly associated with waves breaking approximately symmetrically about the mean downwind direction. In weaker winds the symmetry is not so strong. The symmetry of the momentum flux distribution implies that the net momentum flux is approximately in the wind direction.

The existence of an approximate equilibrium subrange³ in the spectrum of the surface wave field would be of considerable importance for describing the surface wave field and related air-sea fluxes. These measurements offer some insight into this question. Our measurements of $\Lambda(c) \propto c^{-6}$, and $\psi(k_1) \propto k_1^{-5/2}$, are quantitatively consistent with the equilibrium theory of Phillips³ for a subrange of

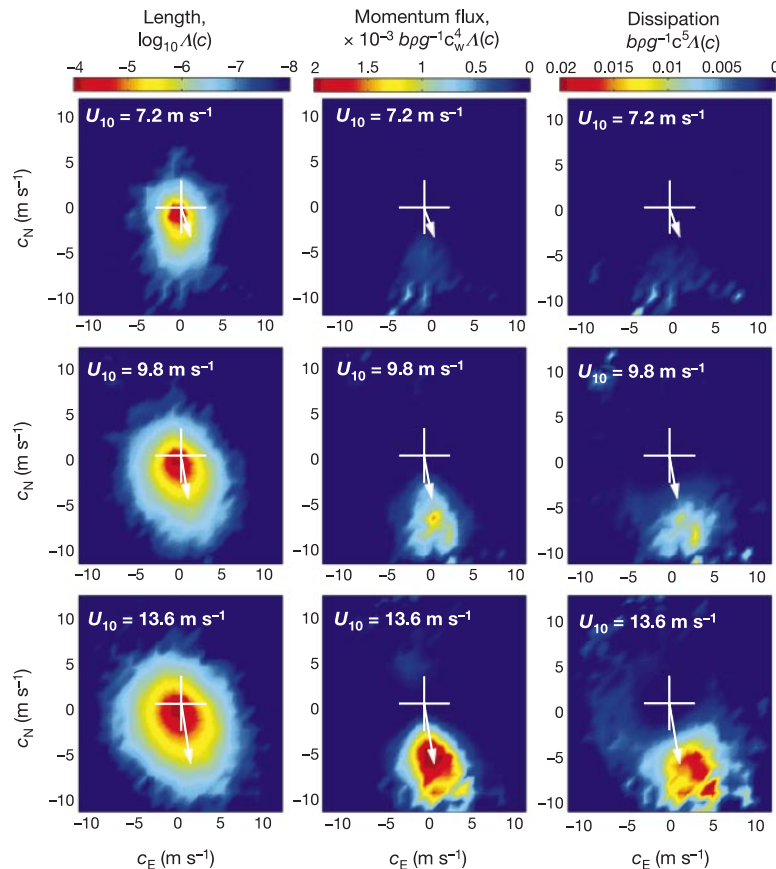


Figure 4 Directional distributions of $\Lambda(c)$ and its moments. **a**, Distributions with respect to the northerly and easterly components of $c(c_N, c_E)$ of $\Lambda(c)$ along with the average wind direction (arrow). The dominant orientation is close to the wind direction. However, there is some upwind and cross-wind breaking (this is accentuated by the logarithmic colour scale of the amplitude). The extremes of these distributions account for the few outliers at

speeds greater than the mean wind speeds in Fig. 2a, e. **b**, The corresponding momentum flux from waves to currents due to breaking for the three wind speeds. As the wind increases the region of significant momentum flux becomes approximately symmetrical about the downwind direction (compare Fig. 2b). **c**, The corresponding distribution of the dissipation with (c_N, c_E) .

larger c and smaller k values, respectively. However, the c^{-6} behaviour is only a local approximation to the exponential distribution.

Hara and Belcher¹⁶ predict an equilibrium range bounded by two asymptotic regimes separated by a 'sheltering wavenumber', k_s . For $k \ll k_s$, they support Phillips' predictions. At larger wavenumbers, $k > k_s$, sheltering by the longer waves produces less growth as momentum flux is lost to the longer waves. The significant decrease in the dissipation that we find for smaller c and larger k are qualitatively consistent with their model, as are the measured one-dimensional surface wavenumber spectra (see also Banner *et al.*¹⁸); however, important differences remain since we find the omni-directional dissipation $D(k)$ decaying as $k^{-7/2}$ compared to their prediction of $k^{-5/2}$. Furthermore, their predictions of k_s for our range of wind speeds are all greater than 100 rad m^{-1} , well outside the range of our data and above their estimate of the high wavenumber cut-off for surface waves^{19,20}.

In estimating the dissipation from the measurements of $\Lambda(c)$, we have assumed that the numerical parameter, b , is effectively constant; this is consistent with ref. 3, and our measurements of $\Lambda(c)$ over a limited range of c . However, laboratory measurements⁵ show that b increases approximately linearly with wave slope, or with breaking going from spilling to plunging. A decrease in b with increasing c would lead to a small decrease in the dissipation.

The exponential distribution of $\Lambda(c)$ is consistent with the fact that Λ must remain bounded for small c , because there cannot be an infinite length of orthogonally finite structures in a finite area. Hence, $c\Lambda(c) \rightarrow 0$ as $c \rightarrow 0$, and there must be a maximum in $c\Lambda(c)$ for small c ; although the data do not clearly represent it. The e-folding scale, found here to be constant, is expected to be a function of the fetch. With the shift in the peak of the wave spectrum to lower frequencies (higher c) as the fetch increases, we expect $\Lambda(c)$, and hence the e-folding scale, to increase; that is, there will be more breaking at larger c .

These measurements are limited to whitecaps with significant air entrainment. Breaking gravity-capillary waves near the minimum phase speed (23 cm s^{-1}), which are not measured by these techniques, may display different behaviour. Thermal imagery, rather than visible imagery, is likely to be important for resolving the statistics of microscale (gravity-capillary) breaking².

This work has demonstrated that breaking, a process vital to many aspects of air-sea interaction, can be measured by remote techniques, and that $\Lambda(c)$ and its moments, which are directly relevant for ocean-surface processes, can be simply characterized. We find that the momentum flux to currents, and wave dissipation, are proportional to U_{10}^3 and dominated by intermediate scale waves. The fraction of the ocean surface mixed ('renewed') by breaking waves per unit time $c\Lambda(c)$, an important component of air-sea heat and gas flux models, is proportional to U_{10}^3 and is dominated by the slower (shorter) breaking waves. Marine spray/aerosol generation is poorly understood with model predictions varying by orders of magnitude²¹. However, the direct generation of spray by breaking fronts is simply proportional to $\Lambda(c)$, albeit with additional dependences on c and the Weber number. Finally, recent research demonstrates that the optical properties of ocean surface waters are influenced by bubbles entrained by breaking waves²², so that accurate optical remote sensing of ocean productivity (such as with chlorophyll a) must account for the statistics of breaking as described here. □

Methods

The video camera (Pulnix TM-9700, 484×768 pixels), attached to a six-degrees-of-freedom motion package (Watson BA-604), was suspended in a pod beneath the Long EZ aircraft. Images were acquired and stored at 5 Hz. The data/image acquisition computer and GPS receiver and antenna were housed under the aircraft's canopy. A base station provided differential-GPS location of the aircraft at 10 Hz. The aircraft's nominal altitude for image-acquisition flights was 450 m, giving pixel sizes of approximately 50 cm.

On the basis of the brightness distribution of each image, a threshold was chosen to distinguish breaking waves and their boundaries from the background. A range of

thresholds was tested to ensure insensitivity of the results to the specific choice. Only actively breaking (expanding) whitecaps were considered. Square sub-images (5×5 pixels on a side) were used for particle imaging velocimetry (PIV) analysis¹⁵ of the velocity of the whitecap boundary. With a 5×5 window the velocity of the boundary was calculated over a 2.5-m segment, approximately every 0.5 m along the boundary: a running average. PIV processing resolves the peak in the correlation to within half a pixel, or 0.3 m normal to the boundary.

In interpreting the whitecap motion, we need to avoid two effects: Doppler shifting (advection); and straining, due to longer waves and swell. In the laboratory^{6,7}, the upstream boundary of the whitecap remains effectively stationary in the absence of the orbital motion of the following waves, while the breaking front moves forward at approximately 80% of the phase speed. In the ocean the advection of the whitecap by longer waves cannot be neglected and so we choose to remove this effect by defining the speed of advance of the whitecap relative to the upstream, or rearward, edge of the whitecap. To avoid these effects we process the data in the following way: (1) a mean velocity is calculated for each whitecap from all the velocities of the boundary elements; (2) the rear of the whitecap is defined as the points on the contour located within an angle of $\pm 60^\circ$ from the centroid of the whitecap in the direction opposite to that of the mean whitecap velocity; (3) the mean velocity of the rear of the whitecap is subtracted from all velocity vectors in the whitecap profile; (4) the components of velocity normal to the whitecap boundary are calculated; (5) the front of the whitecap is defined as the points on the contour located within an angle of $\pm 90^\circ$ from the centroid in the direction of the mean velocity of the whitecap; (6) the statistical analysis includes only elements of the contour belonging to the front of the whitecap; and (7) all elements of the processing were tested for sensitivity to specific choices of thresholds.

To summarize, $\Lambda(c)dc = (1/A_s)\sum_{\substack{c_{\perp i} \in [c, c+dc] \\ A_i}} A_i dc$, where A_i are lengths of the whitecap boundary elements which lie within $\pm 90^\circ$ of the whitecap mean velocity relative to the centroid (calculated for each whitecap), and $c_{\perp i}$ is the outward velocity component normal to A_i . The summation is performed for all whitecap data within the imaged (swept) area A_s .

The significance of the straining of the whitecap by the orbital motion of the longer waves can be estimated by accounting for the relative motion between the forward and rearward faces of the whitecap. If the speed of advance of the whitecap scales with the phase speed of the breaking wave, c_b , and if the length of the whitecap is at most the wavelength of the breaking wave, the relative error in c is $O(a_1 k_1) c_b/c_1$, where $a_1 k_1$ is the slope of the longer wave and c_1 its phase speed. If both $a_1 k_1$ and c_b/c_1 are much less than unity, the error will be very small. As c_b/c_1 tends to unity, the 'long' wave becomes the breaking wave and the laboratory data⁶ show that the issue of straining becomes moot.

One-dimensional wavenumber spectra of surface waves were measured using surface displacement measurements by three Riegl laser altimeters (corrected by a GPS-based aircraft motion/altitude system) during the upwind and downwind low altitude flight legs along the same track as the image data were acquired. In computing the wave spectra the following assumptions have been made: (1) the wind waves are predominantly propagating in the direction of the wind; (2) the statistical properties of the surface are stationary; and (3) the flight track was almost straight with transverse aircraft speeds small compared to the along-track component.

We can show that, using the speed over the ground to transform the time series of sea surface elevation to spatial series and averaging, the (perceived) upwind and downwind wavenumber spectra lead to errors in the wavenumber of $O(cU_a/V_a^2)$, where c is the phase speed of the waves, U_a the wind speed at the aircraft elevation, and V_a the aircraft speed. For these experiments this error is in the range of 0.3–6%.

Using the aircraft position/speed data, the surface elevation data were interpolated to a regular sampling interval in space. Spectra were calculated within a running 1,024-point window (~ 20 s, ~ 1 km), and averaged over the track length of ~ 20 km.

All data presented here were collected during weather events with winds directed nearly parallel to the shoreline (within $5\text{--}10^\circ$). The wind-wave fetch was in the range 100–150 km. The data included in the analysis were taken no closer than 30 km to shore.

Phillips³ predicted that $\Lambda(c) \propto u_*^2 c^{-6}$, where u_* is the friction velocity in the air. Here we scale Λ with U_{10}^3 , since over the range of conditions of these experiments we expect the drag coefficient $C_d \equiv (u_*^2/U_{10}^2)$ to be constant to within a factor of $O(1)$.

If the waves are sufficiently long and steep, dissipation in the bottom boundary layer may need to be considered in interpreting data at large values of c . Estimates of bottom friction are site-specific and based on semi-empirical unsteady boundary-layer theory. Using recognized methods^{23,24} and a range of bottom velocities, we find that the characteristic time for wave decay due to bottom friction in 20–25 m depth is typically $O(10^4)$ seconds, compared to $O(10^3)$ seconds, owing to breaking. Thus the contribution of bottom dissipation to the energy budget is small compared to dissipation due to breaking. Furthermore, measurements of $\Lambda(c)$ in 40–50 m depth were almost indistinguishable from those in 20–25 m.

Received 28 August 2001; accepted 20 February 2002.

- Melville, W. K. The role of surface wave breaking in air-sea interaction. *Annu. Rev. Fluid Mech.* **28**, 279–321 (1996).
- Jessup, A. T., Zappa, C. J., Loewen, M. R. & Healy, V. Infrared remote sensing of breaking waves. *Nature* **385**, 52–55 (1997).
- Phillips, O. M. Spectral and statistical properties of the equilibrium range in wind-generated gravity waves. *J. Fluid Mech.* **156**, 505–531 (1985).
- Duncan, J. H. An experimental investigation of breaking waves produced by a towed hydrofoil. *Proc. R. Soc. Lond. A* **377**, 331–348 (1981).
- Melville, W. K. Energy dissipation by breaking waves. *J. Phys. Oceanogr.* **24**, 2041–2049 (1994).
- Rapp, R. J. & Melville, W. K. Laboratory measurements of deep-water breaking waves. *Phil. Trans. R. Soc. Lond. A* **331**, 735–800 (1990).
- Melville, W. K. & Rapp, R. J. Momentum flux in breaking waves. *Nature* **317**, 514–516 (1985).

8. Loewen, M. A. & Melville, W. K. Microwave backscatter and acoustic radiation from breaking waves. *J. Fluid Mech.* **224**, 601–623 (1991).
9. Phillips, O. M., Posner, F. L. & Hansen, J. P. High range resolution radar measurements of the speed distribution of breaking events in wind-generated ocean waves: Surface impulse and wave energy dissipation rates. *J. Phys. Oceanogr.* **31**, 450–460 (2001).
10. Monahan, E. C. & Muirchearthaigh, I. Optimal power-law description of oceanic whitecap coverage dependence on wind speed. *J. Phys. Oceanogr.* **10**, 2094–2099 (1980).
11. Wu, J. Variations of whitecap coverage with wind stress and water temperature. *J. Phys. Oceanogr.* **18**, 1448–1453 (1988).
12. Donelan, M. A., Longuet-Higgins, M. S. & Turner, S. Periodicity in whitecaps. *Nature* **239**, 449–451 (1972).
13. Lamarre, E. & Melville, W. K. Air entrainment and dissipation in breaking waves. *Nature* **351**, 469–471 (1991).
14. Lamarre, E. & Melville, W. K. Void-fraction measurements and sound speed fields in bubble plumes generated by breaking waves. *J. Acoust. Soc. Am.* **95**, 1317–1328 (1994).
15. Willert, C. & Gharib, M. Digital particle image velocimetry. *Exp. Fluids* **10**, 181–193 (1991).
16. Hara, T. & Belcher, S. E. Wind forcing in the equilibrium range of wind-wave spectra. *J. Fluid Mech.* (submitted).
17. Phillips, O. M. *The Dynamics of the Upper Ocean* (Cambridge Univ. Press, Cambridge, 1977).
18. Banner, M. L., Jones, I. S. F. & Trinder, J. C. Wavenumber spectra of short gravity waves. *J. Fluid Mech.* **198**, 321–344 (1989).
19. Jähne, B. K. & Riemer, K. S. Two-dimensional wavenumber spectra of smallscale water surface waves. *J. Geophys. Res.* **95**, 11531–11546 (1990).
20. Fedorov, A. V., Melville, W. K. & Rozenberg, A. An experimental and numerical study of parasitic capillary waves. *Phys. Fluids* **10**, 1315–1323 (1998).
21. Andreas, E. L. A new sea spray generation function for wind speeds up to 32 m s^{-1} . *J. Phys. Oceanogr.* **28**, 2175–2184 (1998).
22. Terrill, E. J., Melville, W. K. & Stramski, D. Bubble entrainment by breaking waves and their influence on optical scattering in the upper ocean. *J. Geophys. Res.* **106** (C8), 16815–16824 (2001).
23. Madsen, O. S., Poon, Y.-K. & Graber, H. C. in *Coastal Engineering 1988, Proc. 21st Int. Conf. Coastal Eng.* (ed. Edge, W. L.) 492–504 (American Society of Civil Engineers, New York, 1989).
24. Ardhuin, F., Herbers, T. H. C. & O'Reilly, W. C. A hybrid Eulerian-Lagrangian model for spectral wave evolution with application to bottom friction on the continental shelf. *J. Phys. Oceanogr.* **31**, 1498–1516 (2001).

Acknowledgements

These measurements would not have been possible without the fearless flying of T. Crawford and the support of his group at NOAA who made additional airborne measurements. We are grateful to D. Shear for his work on the initial design, construction and testing of the aerial imaging system. W.K.M. thanks T. Hara and O. Phillips for stimulating conversations on the measurement and modelling of equilibrium wave spectra, and O. Madsen and T. Herbers for advice on wave dissipation in bottom boundary layers. We are grateful to F. Veron for carefully reading and commenting on the manuscript. W.K.M. acknowledges the hospitality of the Isaac Newton Institute, Cambridge University, during the revision of this paper. This work was supported by grants from the ONR (Physical Oceanography) and the NSF (Ocean Sciences).

Correspondence and requests for materials should be addressed to W.K.M. (e-mail: kmelville@ucsd.edu).

A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont

Harald Huber*, Michael J. Hohn*, Reinhard Rachel*, Tanja Fuchs*†, Verena C. Wimmer‡ & Karl O. Stetter*

* Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany

‡ Max Planck Institute for Medical Research, Department of Cell Physiology, Jahnstrasse 29, 69120 Heidelberg, Germany

According to small subunit ribosomal RNA (ss rRNA) sequence comparisons all known Archaea belong to the phyla Crenarchaeota, Euryarchaeota, and—indicated only by environmental DNA sequences—to the 'Korarchaeota'^{1,2}. Here we report the cultivation of a new nanosized hyperthermophilic archaeon from

a submarine hot vent. This archaeon cannot be attached to one of these groups and therefore must represent an unknown phylum which we name 'Nanoarchaeota' and species, which we name 'Nanoarchaeum equitans'. Cells of 'N. equitans' are spherical, and only about 400 nm in diameter. They grow attached to the surface of a specific archaeal host, a new member of the genus *Ignicoccus*³. The distribution of the 'Nanoarchaeota' is so far unknown. Owing to their unusual ss rRNA sequence, members

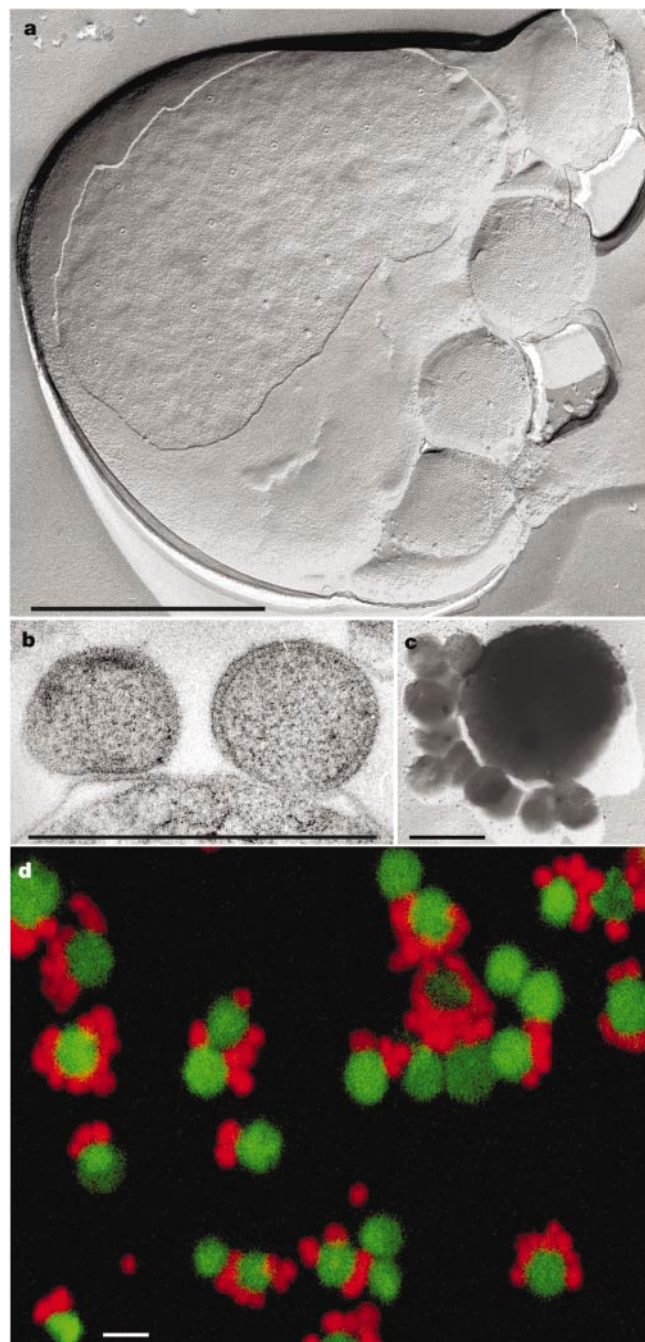


Figure 1 Electron microscopy and fluorescence light microscopy of the 'Nanoarchaeum equitans'–*Ignicoccus* sp. coculture. **a**, Freeze-etched cell of *Ignicoccus* and four attached cells of 'Nanoarchaeum', showing their crystalline S-layer with sixfold symmetry. **b**, Ultrathin section of two cells of 'Nanoarchaeum' attached to the outer membrane of *Ignicoccus*. **c**, Cell of *Ignicoccus*, with several cells of 'Nanoarchaeum' attached on the left side; platinum-shadowed. **d**, Confocal laser scanning micrograph after hybridization with the CY3-labelled probe 515mcR ('Nanoarchaeum') and rhodamine-green-labelled probe CREN499R (*Ignicoccus*). **a–d**, Scale bar, 1.0 μm .

† Present address: AstraZeneca GmbH, Tinsdaler Weg 183, D-22876 Wedel, Germany.

remained undetectable by commonly used ecological studies based on the polymerase chain reaction⁴. '*N. equitans*' harbours the smallest archaeal genome; it is only 0.5 megabases in size. This organism will provide insight into the evolution of thermophily, of tiny genomes and of interspecies communication.

Hyperthermophilic microorganisms with optimal growth temperatures above 80 °C occur in both prokaryotic domains, the Bacteria and Archaea. In phylogenetic trees based on ss rRNA they represent all the short deep lineages, forming a cluster around the root⁵. This feature and their ancient biotopes indicate that hyperthermophiles are still primitive⁵. Within the Archaea, hyperthermophilic members of the Crenarchaeota and the Euryarchaeota have been cultivated, which exhibit unusual cell morphologies such as disks, networks, 'golf clubs', irregular spheres, and branched and irregular rods⁶. Variations in cell volume by up to four orders of magnitude even within a pure culture were observed (for example, *Thermoproteales* and *Desulfurococcales*)⁶.

To investigate hot submarine vent microbial communities, we carried out experiments to cultivate hyperthermophiles from samples of originally hot rocks and gravel taken at the Kolbeinsey ridge, north of Iceland⁷. By anaerobic incubation at 90 °C in the presence of S, H₂ and CO₂, a new autotrophic sulphur-reducing species of the archaeal genus *Ignicoccus*³ could be enriched. In contrast to known species, several large spherical *Ignicoccus* cells were covered by very tiny cocci which could be stained by DNA-specific fluorescence microscopy (DAPI)⁸. Few such tiny cocci existed in the free state.

These tiny cocci could be physically isolated either by using 'optical tweezers'^{9,10} or by ultrafiltration (pore width, 0.45 µm). All attempts to grow them in pure culture employing various inorganic and organic energy sources failed. However, isolation of a combination of a tiny coccus attached to an *Ignicoccus* sphere after incubation under enrichment conditions, resulted in a defined coculture in which about half of the *Ignicoccus* cells appeared to be colonized by the tiny cocci. The final density of both the tiny cocci and the *Ignicoccus* cells was about 3 × 10⁷ cells ml⁻¹ resulting in about two tiny cocci per *Ignicoccus* cell on average. The purified coculture was used in all further investigations. Cloning of single *Ignicoccus* cells gave rise to cultures which never contained tiny cocci.

By electron microscopy, a close attachment of the tiny cocci to the surface of the *Ignicoccus* cells became evident (Fig. 1a, b, c). The tiny cocci consistently exhibited a cell diameter of about 400 nm. In contrast to *Ignicoccus*, they were covered by a regular surface layer (S-layer) with sixfold symmetry and a lattice constant of 15 nm (Fig. 1a). Ultrathin sections showed the presence of cytoplasmic

membranes inside the tiny cocci and the *Ignicoccus* cells (Fig. 1b). No specific attachment structures could be detected thus far at the sites of contact which suggested a loose connection of the tiny cocci with the *Ignicoccus* cells. In line with this assumption, the tiny cocci could be removed by mild sonification (30 W).

The phylogenetic relationships of the tiny cocci were investigated by ss rRNA sequence comparisons^{1,11}. Total DNA of the coculture was extracted and ss rRNA genes were PCR-amplified using primers addressed to highly conserved gene sections. Surprisingly, however, only the *Ignicoccus* ss rRNA gene sequence was amplified even when primers considered general for all Archaea as well as for all organisms ('universal') had been employed (for example, 8aF¹² and 1513uR¹²). More directly, we examined the coculture-derived DNA for ss rRNA genes by Southern blot hybridization¹³, taking advantage of the generally high sequence homology of all ss rRNA genes. Two different hybridization signals became visible, indicating two non-identical ss rRNA genes (Fig. 2; lanes 2 and 4). As a control, DNA isolated from the separate *Ignicoccus* sp. by the same procedure yielded only one hybridization signal (Fig. 2; lanes 1 and 3) which was identical with one from the mixed culture. Therefore, the tiny cocci must have given rise to the second hybridization signal, indicating that they hold a ss rRNA gene different from *Ignicoccus*.

The corresponding DNA fragment was cloned and sequenced. Its sequence turned out to be (so far) unique, harbouring many base exchanges even in the so-called 'highly conserved regions' that are usually employed as primer targets for ss rDNA PCR (Table 1). This explains our initial failure to amplify this gene by PCR. Comparison of the new ss rRNA sequence with those of other microorganisms established the tiny cocci as previously unknown members of the archaeal domain indicated by sequence identities of 0.69 to 0.81 with the Archaea in contrast to sequence identities of only 0.59 to 0.70 with bacterial species. The secondary structure of the ss rRNA sequence conformed to the standard two-dimensional structural model¹¹. It exhibited features characteristic of Archaea, especially helices numbers 17, 18, 36 and 47 (ref. 11; Fig. 3). The sequence identities to the Crenarchaeota (0.73–0.81), the Euryarchaeota (0.69–0.81), and the 'Korarchaeota' (0.73–0.75) were in the same range as those between these three known phyla of Archaea (0.69–0.83). Therefore, the tiny cocci represent a new archaeal phylum. On the basis of its extremely small cell size, we name it 'Nanoarchaeota' (the dwarf archaea) and the corresponding species '*Nanoarchaeum equitans*' (riding the fire sphere).

In ss rRNA phylogenetic trees that are founded on the maximum

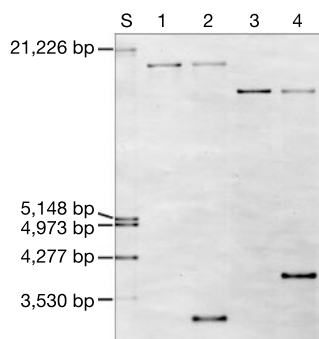


Figure 2 Southern blot analysis of DNA from *Ignicoccus* sp. and the '*N. equitans*'–*Ignicoccus* sp. coculture, treated with restriction enzymes. S, length standard (DNA relative molecular mass marker III, Roche). Lane 1 and 3, *Ignicoccus* DNA. Lane 2 and 4, DNA of the coculture. Lane 1 and 2, digested with *Eco*RI. Lane 3 and 4, digested with *Hind*III.

Table 1 Sequence specificity of ss rDNA primers

Designation	Sequence (5' → 3')	Specific against				
		C	E	K	N	B
8aF ¹²	TCY GGT TGA TCC TGC C	+	+	+	–	–
8mcF	TCC <u>CGT</u> TGA TCC TGC <u>G</u>	–	–	–	+	–
9bF ¹²	GRG TTT GAT CCT GGC TCA G	–	–	–	–	+
	<u>CCC</u> GTT GAT CCT <u>GCG</u> GGA G	–	–	–	+	–
217kF ²⁹	GAG GCC CCA GGR TGG GAC CG	–	–	+	–	–
	<u>CCC</u> GCC CGA GGA TGG <u>GCC</u> <u>GG</u>	–	–	–	+	–
519uF ¹²	CAG CMG CCG CGG TAA TAC	+	+	–	–	+
519mcF	CAG CCG CCG CGG <u>GAA</u> <u>CAC</u>	–	–	–	+	–
1114aR	GGG TCT CGC TCG TTR CC	+	+	+	–	–
1114mcR	GGG TCT CGC <u>CTG</u> TTT CC	–	–	–	+	–
1406uR ¹²	ACG GGC GGT GTG TRC AA	+	+	+	–	+
1406mcR	ACG GGC GGT <u>GAG</u> TGC AA	–	–	–	+	–
1513uR ¹²	ACG GHT ACC TTG TTA CGA CTT	+	+	+	–	+
1513mcR	ACG GCT ACC TTG <u>TGT</u> CGA CTT	–	–	–	+	–
EURY498R ¹⁵	CTT GCC CRG CCC TT	–	+	–	–	–
511mcR	CTT GCC CAC CGC TT	–	–	–	+	–
CREN499R ¹⁵	CCA GWC TTG CCC CCC GCT	+	–	–	–	–
515mcR	CCC CTC TTG CCC ACC GCT	–	–	–	+	–
ARCH915R ¹⁶	GTG CTC CCC CGC CAA TTC CT	+	+	–	–	–
934mcR	GTG CTC CCC CGC <u>CTA</u> TTC CT	–	–	–	+	–

C, Crenarchaeota; E, Euryarchaeota; K, 'Korarchaeota' sequences; N, 'Nanoarchaeota'; B, Bacteria. Base exchanges in '*N. equitans*' are underlined.

parsimony, distance matrix, and maximum likelihood methods¹⁴, '*N. equitans*' represents an isolated, very deeply branching lineage. However, because of its unique ss rRNA, a large variation of its branching point, challenged by insignificant bootstrap values below 50%, was observed (not shown). Therefore, the determination of the accurate branching position of the 'Nanoarchaeota' must await

the discovery of further members of this group.

Owing to their great divergences in ss rRNA—in contrast to the *Ignicoccus* host—cells of the 'Nanoarchaeota' did not stain by fluorescence *in situ* hybridization using ss rRNA-targeted oligonucleotide probes directed against Crenarchaeota and Euryarchaeota (for example, EURY498R¹⁵, CREN499R¹⁵ and ARCH915R¹⁶;

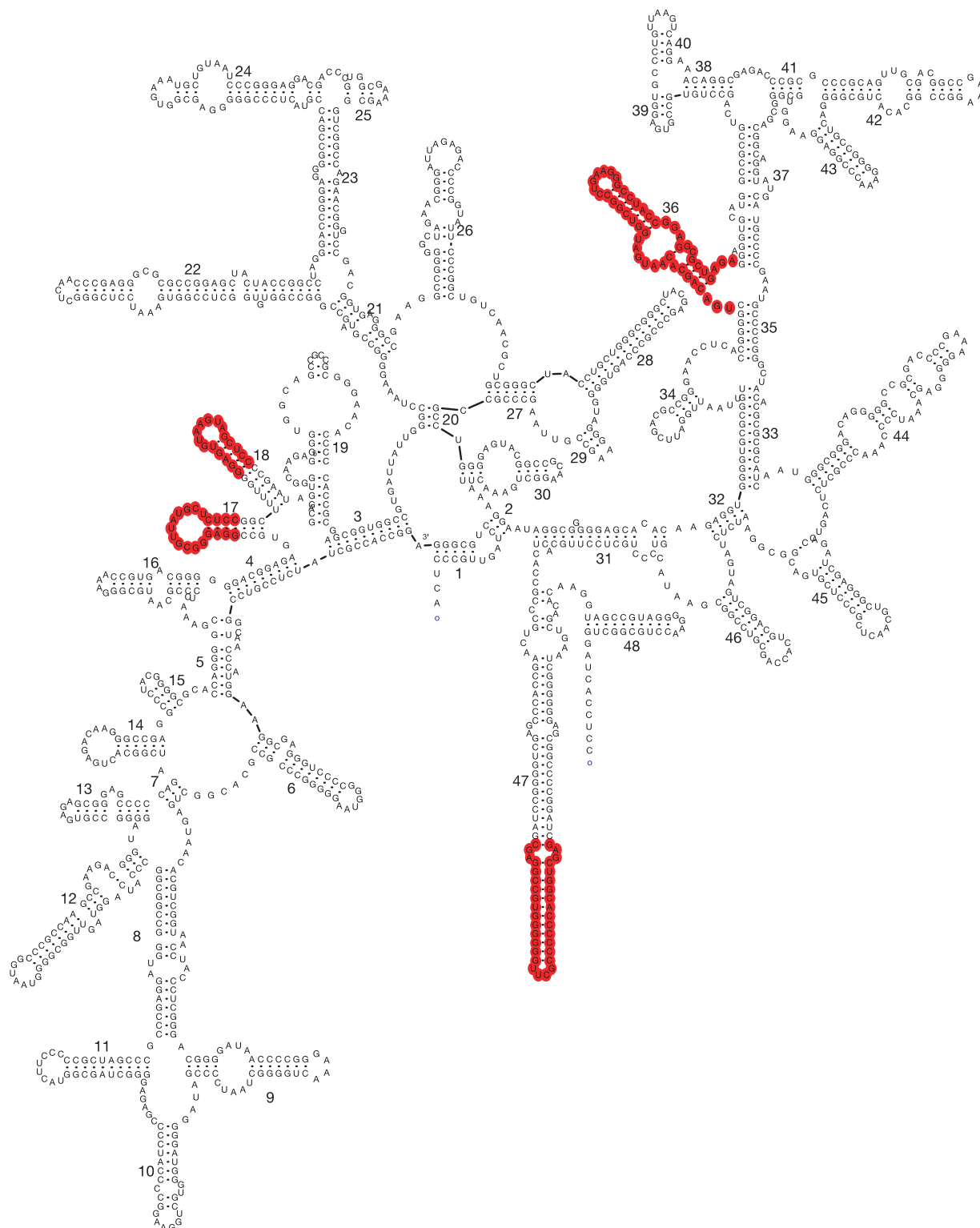


Figure 3 Secondary structure model for the ss rRNA of '*N. equitans*'. Highlighted positions indicate characteristic archaeal secondary structures¹¹. Numbers correspond to the helix numbers. The model structure was determined using RnaViz²⁸.

Table 1). However, after redesigning these oligonucleotide probes on the basis of the 'Nanoarchaeota' sequence (for example, 511mCR, 515mCR and 934mCR; Table 1), the tiny cocci exhibited a bright fluorescence (Fig. 1d), indicating that they possessed ribosomal RNA harbouring the target sequence.

Few physiological properties of the 'Nanoarchaeum' strain are known thus far. Isolated cells did not grow on cell homogenates of *Ignicoccus* sp., so they require an actively growing *Ignicoccus* culture. Within pure cultures of *Ignicoccus*, 'Nanoarchaeum' cells locally separated by dialysis bags were unable to propagate (not shown). Therefore, a direct cell-cell contact with the host appears to be a prerequisite for growth. No negative influence on *Ignicoccus*, such as slower growth, lower final cell density, or increased cell lysis, could be observed in the coculture compared to the pure culture, rendering a parasitic mode of living of 'Nanoarchaeum' unlikely. Therefore, the organism is most probably symbiotic. So far, archaeal symbionts are only known at mesophilic temperatures in association with eukaryotes (see ref. 17 for example) and bacteria¹⁸. Cocultures of 'Nanoarchaeum' grew within the same temperature range (70–98 °C) as its host. During mass-culturing in a 300-litre enamel-protected fermentor³, final cell concentrations of 'Nanoarchaeum' could be improved about tenfold by an increased gassing rate (201 min⁻¹; H₂ : CO₂ = 80 : 20), while that of *Ignicoccus* remained unchanged. This procedure improved hydrogen supply and efficiently removed H₂S, the main metabolic end product of *Ignicoccus*³. During the late exponential growth phase, about 80% of the 'Nanoarchaeum' cells detached from their host cells and occurred freely in suspension. They could be collected (for example, 1.5 g wet weight) by high-speed centrifugation, after removal of the *Ignicoccus* cells at low speed. By this method, cell masses of 'Nanoarchaeum' became available for biochemical and molecular investigations. A first analysis of its genome size by adding up the sizes of restriction fragments determined after pulse field gel electrophoresis resulted in only 500 kilobases (kb), one of the smallest genomes of all prokaryotes known so far (not shown)¹⁹.

With its tiny cell and genome size, 'Nanoarchaeum' resembles an intermediate between the smallest living organisms like *Mycoplasma genitalium*¹⁹ and big viruses like the pox virus and *Chlorella* virus CVK2 (refs 20 and 21), and is close to the theoretical minimum genome size calculated for a living being²². Small cell size and reduced genomes are common features observed in many symbiotic or parasitic Bacteria²³, but have so far not been reported for Archaea or hyperthermophiles. No final conclusions can be drawn about a primitive or derived state of evolution of 'Nanoarchaeum'. However, its high growth temperature and anaerobic mode of life correlates with probable early environmental conditions which suggest that the 'Nanoarchaeota' are possibly still a primitive form of microbial life. Analysing its genome and its symbiotic relationship should help. Other members of the 'Nanoarchaeota' may still exist on Earth and may be discovered by using our knowledge of their unique ss rRNA. □

Methods

Cultivation conditions

Cultures were grown anaerobically in serum bottles as described³. Mass culturing was carried out in the same medium in a 300-litre enamel-protected fermentor (gassing rate routinely 21 min⁻¹; H₂ : CO₂ = 80 : 20). Cell mixtures were collected by centrifugation (9,000 r.p.m.; 30 min using rotor GS3 (Sorvall)) and resuspended in sulphur-free culture medium. The *Ignicoccus* cells were removed by centrifugation (2,000 r.p.m. for 30 min using rotor SS34 (Sorvall)). Then the 'Nanoarchaeum' cells were precipitated (15,000 r.p.m. for 15 min using rotor SS34). Axenic cultivation of 'Nanoarchaeum equitans' was tested autotrophically in the presence of H₂ : CO₂ (80 : 20) and elemental sulphur, nitrate, nitrite, and sulphate (each 0.1% w/v) as possible electron acceptors, and heterotrophically on single and complex organic nutrients such as sugars, amino acids, yeast extract, peptone, bacterial and archaeal (for example, from *Ignicoccus*) cell extracts (each 0.1% w/v; gas phase: H₂ : CO₂ = 80 : 20 and N₂ : CO₂ = 80 : 20).

Light and electron microscopy

Phase-contrast and electron microscopy were carried out as described³. Confocal image series were recorded on a LEICA TCS SP2 laser scanning microscope using the laser wavelengths 488 nm (argon) and 543 nm (HeNe) excitation. Fluorescence *in situ* hybridization (FISH) was performed as described¹⁵. CY3- and rhodamine-green-labelled oligonucleotides were obtained from Metabion (Germany).

Southern blot experiments

DNAs were digested with restriction endonucleases *EcoRI* and *HindIII*. The fragments were separated by electrophoresis (1% Seakem ME agarose) and transferred to a positively charged nylon membrane by Southern blot¹³. The membrane was hybridized with a digoxigenine-labelled single-stranded DNA fragment of an archaeal ss rRNA gene (*Metallosphaera sedula*; length 600 nucleotides, position 519 to 1119, *Escherichia coli* numbering²⁴).

Estimation of the genome size

For pulse field gel electrophoresis 'Nanoarchaeum' cells were embedded in 100-µl agarose plugs (0.8% Incert agarose, FMC; final concentration: 3 × 10⁸ cells per plug). After cell lysis according to lysis method 'three'²⁵ genomic DNA was digested using the restriction enzymes *AscI*, *BssHII*, *NotI*, or *SacII* (each 100 units, 18 h). The restriction fragments were separated as described²⁵.

Small subunit rDNA sequence analysis

The nearly complete ss rRNA gene of the new *Ignicoccus* species was PCR-amplified and sequenced as described³. To obtain the ss rRNA gene of 'N. equitans', genomic DNA of the coculture was digested with *EcoRI* and separated as described (see 'Southern blot'). DNA fragments of the correct size, identified by Southern blot, were extracted, cloned into a plasmid (pBluescript SK(-)) and amplified in *Escherichia coli* (strain Dh5α). These inserts were sequenced by primer walking using the primers 1406uR¹², 934mCR, 1060aF²⁶, 515mCR, 519mCF and 8mCF (Table 1). The new sequences were aligned with approximately 10,000 homologous sequences available in public databases using the automatic alignment tools of the ARB package²⁷. Distance matrix (Jukes-Cantor correction), maximum parsimony, and maximum likelihood (fastDNAmI) methods were used for tree reconstruction as implemented in the ARB package. The secondary structure of the ss rRNA of 'N. equitans' was displayed using the program RnaViz²⁸.

Received 7 November 2001; accepted 29 January 2002.

- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Barns, S. M., Delwiche, C. F., Palmers, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
- Huber, H. *et al.* *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species, *Ignicoccus islandicus* sp. nov. and *Ignicoccus pacificus* sp. nov. *Int. J. Syst. Evol. Microbiol.* **50**, 2093–2100 (2000).
- Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl Acad. Sci. USA* **91**, 1609–1613 (1994).
- Stetter, K. O. Microbial life in hyperthermal environments. *Am. Soc. Microbiol. News* **61**, 285–290 (1995).
- Stetter, K. O. *Size Limits of Very Small Microorganisms; Proceedings of a Workshop* (ed. Space studies board) 68–73 (National Academic Press, Washington DC, 1998).
- Fricke, H., Giere, O., Stetter, K. O., Alfredsson, G. A. & Kristjansson, J. K. Hydrothermal vent communities at the shallow subpolar mid-Atlantic ridge. *Mar. Biol.* **102**, 425–429 (1989).
- Huber, H., Huber, G. & Stetter, K. O. A modified 4',6'-diamidino-2-phenylindole fluorescence staining procedure suitable for the visualization of lithotrophic bacteria. *Syst. Appl. Microbiol.* **6**, 105–106 (1985).
- Ashkin, A., Dziedzic, J. M. & Yamane, T. Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**, 769–771 (1987).
- Huber, R. *et al.* Isolation of a hyperthermophilic archaeum predicted by *in situ* RNA analysis. *Nature* **376**, 57–58 (1995).
- Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **51**, 221–271 (1987).
- Eder, W., Ludwig, W. & Huber, R. Novel 16S rRNA gene sequences retrieved from highly saline brine sediments of Kebrit Deep, Red Sea. *Arch. Microbiol.* **172**, 213–218 (1999).
- Sambrook, J. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, New York, 1989).
- Ludwig, W. *et al.* Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**, 554–568 (1998).
- Burggraf, S. *et al.* Identifying members of the domain Archaea with rRNA-targeted oligonucleotide probes. *Appl. Environ. Microbiol.* **60**, 3112–3119 (1994).
- Stahl, D. A. & Amann, R. *Nucleic acid techniques in bacterial systematics* (eds Stackebrandt, E. & Goodfellow, M.) 205–248 (Wiley, Chichester, 1991).
- Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl Acad. Sci. USA* **93**, 6241–6246 (1996).
- Boetius, A. *et al.* A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**, 623–626 (2000).
- Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
- Rohozinski, J., Girtton, L. E. & Van Etten, J. L. *Chlorella* viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. *Virology* **168**, 363–369 (1989).
- Murphy, F. A. *et al.* (eds) *Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses* (Springer, Vienna/New York, 1995).

22. Hutchinson, C. A. III *et al.* Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169 (1999).
23. Ochman, H. & Moran, N. A. Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096–1099 (2001).
24. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J. Mol. Biol.* **148**, 107–127 (1981).
25. Baumann, C., Judex, M., Huber, H. & Wirth, R. Estimation of genome sizes of hyperthermophiles. *Extremophiles* **2**, 101–108 (1998).
26. Burggraf, S., Huber, H. & Stetter, K. O. Reclassification of the crenarchaeal orders and families in accordance with 16S ribosomal RNA sequence data. *Int. J. Syst. Bacteriol.* **47**, 657–660 (1997).
27. Ludwig, W. & Strunk, O. ARB: A software environment for sequence data (2002); available at <http://www.arb-home.de>.
28. De Rijk, P. & De Wachter, R. RnaViz, a program for the visualisation of RNA secondary structure. *Nucleic Acids Res.* **25**, 4679–4684 (1997).
29. Burggraf, S., Heyder, P. & Eis, N. A pivotal archaea group. *Nature* **383**, 780 (1997).

Acknowledgements

We thank W. Ludwig and D. Prangishvili for stimulating discussions, S. Diller, S. Leptihn, M. Brandl, I. Wyszchony and P. Hummel for technical support, and B. Hedlund for critically reading the manuscript. We are grateful to the cruise leader P. Stoffers, the crew of RV *Poseidon* and the submersible *Jago* team for support during sampling, and the Icelandic government for a research permit. This study was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to K.O.S. (e-mail: karl.stetter@biologie.uni-regensburg.de). The 16S rRNA (ss rRNA) sequences for '*Nanoarchaeum equitans*' and *Ignicoccus* sp. were deposited at GenBank under accession numbers AJ318041 and AJ318042, respectively.

Feedback with soil biota contributes to plant rarity and invasiveness in communities

John N. Klironomos

Department of Botany, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Understanding the relative abundance of species in plant communities is an unsolved problem^{1–10}. Mechanisms such as competition, resource partitioning⁵, dispersal ability¹⁰ and predation tolerance^{6–9} do not adequately explain relative abundance under field conditions^{11,12}. Recent work suggests that interactions between plants and soil microbes is important^{13–21}. Here I show that such interaction explains a significant proportion of the variance in the relative abundance of species in plant communities. Rare plants exhibited a relative decrease in growth on 'home' soil in which pathogens had had a chance to accumulate, whereas invasive plants benefited from interactions with mycorrhizal fungi. Some plant species accumulate pathogens quickly and maintain low densities as a result of the accumulation of species-specific pathogens, whereas others accumulate species-specific pathogens more slowly and do not experience negative feedback until plant densities reach high levels^{13,15,21}. These results indicate that plants have different abilities to influence their abundance by changing the structure of their soil communities, and that this is an important regulator of plant community structure.

To test for the possibility that feedback with soil organisms will determine plant abundance, I conducted four separate and complementary experiments. In the first experiment I compared the direction and magnitude of soil feedback in five 'highly invasive'

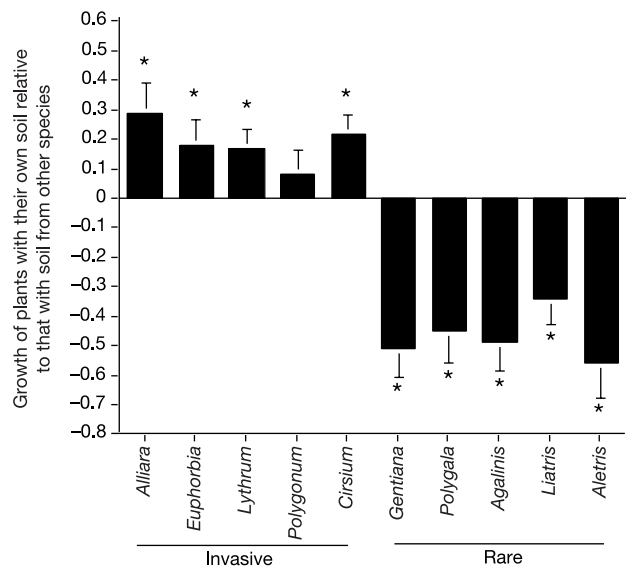


Figure 1 The soil feedback responses of five invasive and five rare plant species. $N = 10$ per treatment. Bars represent the mean with ± 1 s.e. Asterisks represent significant differences in plant growth when grown in home versus foreign soil after t -test analysis.

plant species versus five 'rare and endangered' plants in Canadian old-fields (meadows) and grasslands. I predicted that, in low-density monocultures and under previously uninhabited soil, rare plants will exhibit negative feedback limiting their growth and spread over time, whereas invasive plants will accumulate pathogens more slowly and display neutral or positive feedback. Each plant species was grown for ten weeks from seed in a pot with soil that either had a history of the same plant or a history of a different species of plant (see Methods for more details). Feedback was determined by comparing plant growth responses in their own versus foreign soils.

The results supported the stated prediction. The five rare plant species experienced strong negative feedback when grown in mono-

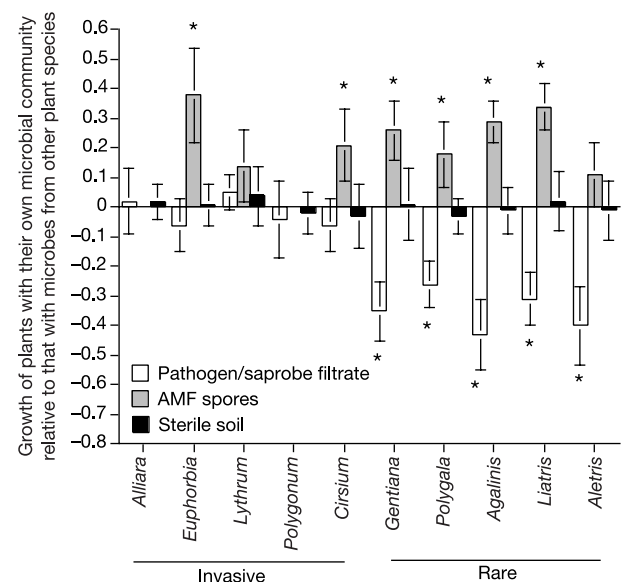
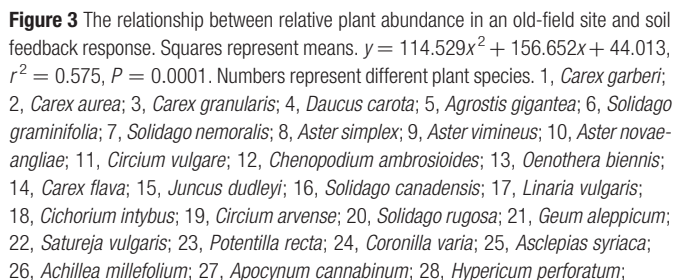


Figure 2 The feedback responses of five invasive and five rare plant species to pathogen/saprobies, mycorrhizal fungi, and sterile soil fractions. Bars represent the mean with ± 1 s.e. Asterisks represent significant differences in plant growth when grown with home versus foreign fractions after t -test analysis.



29, *Agrostis scabra*; 30, *Phleum pratense*; 31, *Poa compressa*; 32, *Echium vulgare*; 33, *Centaurea jacea*; 34, *Rudbeckia serotina*; 35, *Poa pratensis*; 36, *Dactylis glomerata*; 37, *Cerastium vulgatum*; 38, *Galium palustre*; 39, *Oenothera perennis*; 40, *Prunella vulgaris*; 41, *Trifolium pratense*; 42, *Convolvulus arvensis*; 43, *Silene cucubalus*; 44, *Erigeron strigosus*; 45, *Asparagus officinalis*; 46, *Hieracium auranticum*; 47, *Erigeron philadelphicus*; 48, *Veronica officinalis*; 49, *Plantago lanceolata*; 50, *Galium mollugo*; 51, *Hieracium pilosella*; 52, *Vicia cracca*; 53, *Hieracium pratense*; 54, *Medicago lupulina*; 55, *Ranunculus acris*; 56, *Taraxacum officinale*; 57, *Fragaria virginiana*; 58, *Chrysanthemum leucanthemum*; 59, *Tragopogon pratensis*; 60, *Bromus inermis*; 61, *Panicum lanuginosum*.

Two more experiments were set up to test for the possibility that pathogens and mycorrhizal fungi were the agents responsible for the observed feedback responses in the first experiment. In experiment 2, instead of using the entire soil, only specific microbial fractions were used to re-run the original experiment. The fractions were: (1) spores of arbuscular mycorrhizal fungi (AMF); (2) pathogenic/saprobic microbial filtrate; and (3) sterile soil. Once again, the same

ten plant species were grown for ten weeks from seed with one of the soil fractions. The fractions originated from soil containing either a similar or a different plant species (as in experiment 1). Overall, rare plants showed strong negative feedback for the pathogen/saprobe filtrate when grown in soil with a history of the same plant species (Fig. 2). In contrast, the origin of the pathogen/saprobe filtrate did not significantly affect growth of the invasive plant species. Effects of AMF, however, did not differ between invasive and rare plants. For both groups of plants, AMF isolated from soil with a history of the same plant species had a more positive effect on plant growth than fungi isolated from a different plant host. There was no significant feedback effect using the sterile soil, suggesting that the effects are biological rather than chemical. A third experiment was conducted to confirm that pathogenic or parasitic organisms were responsible for the negative feedback responses. The three most commonly

	Invasive					Rare and endangered				
Plant species tested	Ap	Ee	Lsa	Pc	Ca	Ga	Pi	Ag	Lsp	Af
Ap	7	3	-3	-1	3	-5	1	15	-9	-2
Ee	7	-8	-10	9	9	14	5	10	-6	4
Lsa	-8	-6	-6	6	16	-13	-18*	-14	5	-8
Pc	13	5	-15	2	-10	4	7	6	-4	10
Ca	16	9	17	12	7	-15	-8	-2	9	-4
Ga	11	-6	-3	16	-18	-54*	5	7	10	-7
Pi	-14	2	-4	12	7	5	-28*	12	-7	-11
Ag	9	-8	-7	2	-19*	3	-5	-46*	11	-16
Lsp	3	-11	-15	7	3	12	10	-2	-51*	9
Al	-1	5	11	12	-24*	-45*	-3	-5	17	-63*

68  © 2002 Macmillan Magazines Ltd NATURE | VOL 417 | 2 MAY 2002 | www.nature.com

isolated non-AMF fungi from the rhizosphere of each plant species were tested for their effects on the growth of each plant species (Table 1). With all five rare plants, significant reductions in growth were observed when they were inoculated with fungi isolated from their own root systems. The most commonly isolated fungi from these plants were species of *Verticillium*, *Fusarium* and *Cylindrocarpon* and, in all cases, they formed systemic infections in the roots of these plants when re-inoculated. No such growth reductions were observed for any of the invasive plants tested. With both groups of plants, inoculations using fungi from foreign plants rarely resulted in significant growth depressions (Table 1). Overall, these combined results support the hypothesis that feedback between plants and soil communities may strongly determine the ability of a plant to establish, invade and persist in a local habitat. Most plants can potentially experience positive feedback with AMF communities, but this is not realized unless negative feedback with pathogens is lacking.

The first three experiments indicate that soil feedback could be an important mechanism for coexistence and the regulation of plant biodiversity in communities. In the fourth experiment I tested the hypothesis that the direction and degree of soil feedback can account for a large proportion of the variation in relative abundance among species that coexist within a plant community. To test for this I measured soil feedback and relative abundance of each of 61 coexisting plant species within an old-field meadow community in southern Ontario, Canada. Relative plant abundance in the field ranged from 1–81%. Also, there was a wide range of feedback responses, from –0.5 to 0.2. The majority of plants were found to experience a net negative feedback, whereas net positive feedback was found in only 18% of plant species. As predicted, there was a strong positive relationship between soil feedback and relative abundance in the field (Fig. 3). All plant species that were found to have strong negative feedback were also found to have low relative abundance. Conversely, plants with high abundance had either low negative feedback or positive feedback. Variation in degree of feedback accounted for a large proportion (57.5%) of the variation in relative abundance.

The results indicate that soil organisms and their feedback effects on plants can strongly influence the relative abundance of plant species within a community. When introduced to previously uncultured soil at low densities, plants that ultimately achieve high abundance do not seem to accumulate species-specific pathogens at the same rate as plants that remain in low abundance. Negative feedback responses, particularly reductions in plant growth and spread, are pathogen-density dependent^{13,15,16,22}, so plants that accumulate pathogens at a slower rate can potentially reach higher densities in nature. Plants may accumulate pathogens at a slower rate either because of inherent plant traits²², or because they have escaped their harmful pathogens by invading foreign territory²³. Both are likely explanations for the lack of negative feedback in alien invasive plants in this study. Ultimately, even abundant plants are unlikely to expand indefinitely. Plant-specific pathogen loads are maximized under high plant densities, especially under dense monocultures, eventually incurring negative feedback on abundant plants^{13,15,21,22}. Other above- and belowground organisms across trophic groups can further interact with feedback responses²⁴, and probably explain a large proportion of the remaining variance in relative plant abundance. Understanding the mechanisms responsible for the abundance of organisms may lead to new approaches for the management of ecosystems, in particular, conservation of rare and endangered species, and protection of ecosystems from species invasions. □

Methods

Experiment 1

Soil feedback in rare versus invasive plant species (Fig. 1) was examined as follows: seeds of rare and invasive plant species were collected from various field sites across southern

Ontario and Quebec, Canada. The rare species were: *Agalinis gattingeri* (Small) Small (Gattinger's agalinis), *Aletris farinosa* L. (colicroot), *Gentiana alba* Muhl. (white prairie gentian), *Liatris spicata* (L.) Willd. (dense blazing star) and *Polygala incarnata* L. (pink milkwort). The invasive species were: *Alliaria petiolata* (Bieb.) Cavara and Grande (garlic mustard), *Cirsium arvense* (L.) Scop. (Canada thistle), *Euphorbia esula* L. (leafy spurge), *Lythrum salicaria* L. (purple loosestrife) and *Polygonum cuspidatum* Sieb. & Zucc. (Japanese knotweed). The five rare species are all native to North America. They are considered to be rare within prairie communities, and they are further threatened because of the loss of prairie habitat due to agricultural expansion and residential development in southern Canada. The five invasive plants were all introduced to North America over a century ago from Eurasia. Soil was collected from the Long-Term Mycorrhizal Research Site (LTMRS), located within the University of Guelph Arboretum, Canada. A section of the LTMRS was used that did not contain any of the plant species tested. The experiment had three stages, beginning with 200 pots, growing the same plant species twice in succession (stages 1 and 2) and then dividing among 'home' and 'foreign' treatments (stage 3). In 'home' treatments the five invasive and five rare plants were grown in pots with their own respective histories (10 replicates each). In 'foreign' treatments they were grown in pots with histories of the other plants (10 replicates). The ten foreign replicates were comprised of one replicate with a history of each of the remaining nine plant species tested, as well as one additional species (*Taraxacum officinale* Weber ex Wiggers). Each experimental unit consisted of a single seedling growing in an eight-inch pot containing field-collected loam soil (total N = 83.1 mmol kg⁻¹, total P = 6.2 mmol kg⁻¹, percentage organic matter was 6.1), and placed in a random position on a greenhouse bench. Plants were watered on a weekly basis but did not receive additional nutrients. Each plant was allowed to grow for 10 weeks after which the entire plant was harvested. In the second stage, soil from each pot from the first stage was placed into new pots, a new seedling of the same species was grown for another 10 weeks, and then harvested. In the third stage, soil from each pot from the second stage was placed into new pots, a new seedling was grown for another 10 weeks (either the same or a different plant species), and then harvested. At harvest, plant weight was determined after drying at 60 °C for 48 hours. Only plants from the third stage were used to determine feedback with the following equation (for plant A): biomass of plant A (grown in soil with a history of plant A) - biomass of plant A (grown in soil with a history of other plants). For each plant species, comparisons of plant growth in home versus foreign soil were done with Student *t*-tests.

Experiment 2

Soil feedback in rare versus invasive plant species was determined using AMF, pathogen/saprobe and chemical fractions (Fig. 2). After the second round of harvests from the previous experiment (second stage), two 10-g soil subsamples were removed from each pot. The first soil sample was used to extract (1) AMF communities; and (2) soil pathogen/saprobe communities. AMF spores were collected on a 45-µm mesh after wet-sieving²⁵, and surface sterilized with 10% sodium hypochlorite. Pathogens and other non-mycorrhizal microbes were collected from the filtrate that passed through a 20-µm sieve. The second soil sample was sterilized by γ-irradiation, and used as a control for abiotic effects. To determine AMF, pathogen/saprobe, and abiotic feedback, seedlings of each plant were grown in pots containing sterilized soil and one of the three fractions (AMF, pathogen/saprobe, sterilized soil). Plants were grown for 10 weeks as in experiment 1, and then harvested. After harvest, plant weight was determined after drying at 60 °C for 48 hours. There were 10 replicates for each treatment. Feedback was calculated using the following equations: (1) AMF feedback for plant A = biomass of plant A (grown in soil with AMF of plant A) - biomass of plant A (grown in soil with AMF of other plants); (2) pathogen/saprobe feedback for plant A = biomass of plant A (grown in soil with pathogens/saprobe of plant A) - biomass of plant A (grown in soil with pathogens/saprobe of other plants); (3) Sterile soil feedback for plant A = biomass of plant A (grown in soil with sterile soil of plant A) - biomass of plant A (grown in soil with sterile soil of other plants). For each plant species, comparisons of plant growth in home versus foreign soil fraction were done with Student *t*-tests.

Experiment 3

Growth responses of plants to fungi cultured from home versus foreign rhizospheres (Table 1) were determined as follows. At final harvest of experiment 1, twenty 0.5-cm root fragments were plated on a 2% malt-extract agar medium containing Rose Bengal²⁶. Root fragments were obtained from each experimental unit in the 'home' treatment and were surface sterilized with 0.1% mercuric chloride before plating. Plates were incubated for three weeks at 20 °C, and then all resulting fungal colonies were subcultured and identified²⁷. A new experiment was set up to test the influence of fungi isolated from each of 10 plant species on the growth of each of 10 plant species. Each experimental unit consisted of a single plant seedling growing in an eight-inch pot containing sterilized silica sand and fungal inoculum originating from one of the ten plant species. There were 10 replicates per 100 plant/fungal combinations, as well as ten non-inoculated controls for each plant species, resulting in 1,100 experimental units. Plant seeds were originally germinated on moist sterile filter paper. Before being planted in the pots, the roots of one-week-old seedlings were dipped in spore suspensions of the cultured fungi. Spore suspensions were comprised of 1 × 10⁶ spores of each of the three most common fungal species isolated from the respective plant species. Experimental units were placed in a random position on a greenhouse bench. Plants were watered as needed, and fertilized on a weekly basis with half-strength Hoagland's solution. Plants were grown for 10 weeks as in experiment 1, and then harvested. At harvest, plant weight was determined after drying at 60 °C for 48 hours. Subsamples of roots were stained with Chlorazol Black E, and inspected with differential-interference contrast and incandescent light microscopy for confirmation of fungal infection. Plant biomass in inoculated versus non-inoculated plants were compared using Student *t*-tests.

Experiment 4

Soil feedback and relative abundance of plants in the field (Fig. 3) was measured. The abundance of each plant species was measured at 100 different locations within the LTMRs. Locations were randomly chosen, and the presence of all plant species within a 1-m² quadrat was recorded at each location. Relative abundance for each species was calculated as the percentage of locations containing that species. This was performed in the summers of 1998 and 2000, and results were pooled. Seeds were collected from each of 61 plant species, and feedback response was determined using similar methods as described above in experiment 1. Regression analysis was used to determine the relationship between plant abundance and soil feedback.

Received 6 November 2001; accepted 14 February 2002.

- Pysek, P., Prach, K., Rejmanek, M. & Wade, M. *Plant Invasions: General Aspects and Special Problems* (SPB Academic, Amsterdam, 1995).
- Huston, M. A. *Biological Diversity: The Coexistence of Species on Changing Landscapes* (Cambridge Univ. Press, Cambridge, 1994).
- Ricklefs, R. E. & Schluter, D. *Species Diversity in Ecological Communities* (Univ. Chicago Press, Chicago, 1993).
- Gaston, K. J. *Rarity* (Chapman & Hall, London, 1994).
- Grace, J. B. & Tilman, D. *Perspectives on Plant Competition* (Academic, San Diego, 1990).
- Belsky, A. J. Effects of grazing, competition, disturbance and fire on species composition and diversity in grassland communities. *J. Veget. Sci.* **3**, 187–200 (1992).
- Milton, W. The effect of manuring, grazing and cutting on the yield, botanical and chemical composition of natural hill pastures. *J. Ecol.* **28**, 326–356 (1940).
- Crawley, M. J. in *Insect-Plant Interactions* (ed. Bernays, E. A.) Vol. 1, 45–71 (CRC Press, Boca Raton, 1989).
- Reader, R. J. Relationship between species relative abundance and plant traits for an infertile habitat. *Plant Ecol.* **134**, 43–51 (1998).
- Rabinowitz, D., Rapp, J. K., Cairns, S. & Mayer, M. The persistence of rare prairie grasses in Missouri: environmental variation buffered by reproductive output of sparse species. *Am. Nat.* **134**, 525–544 (1989).
- Tokeshi, M. *Species Coexistence: Ecological and Evolutionary Perspectives* (Blackwell Science, Oxford, 1999).
- Tilman, D. in *Biodiversity and Ecosystem Function* (eds Schulze, E.-D. & Mooney, H. A.) 327–344 (Springer, Berlin, 1994).
- Van der Putten, W. H., Van Dijk, C. & Peters, B. A. M. Plant-specific soil-borne diseases contribute to succession in foredune vegetation. *Nature* **362**, 53–56 (1993).
- Grime, J. P., Mackey, J. M. L., Hillier, S. H. & Read, D. J. Floristic diversity in a model system using experimental microcosms. *Nature* **328**, 420–422 (1987).
- Bever, J. D. Feedback between plants and their soil communities in an old field community. *Ecology* **75**, 1965–1977 (1994).
- Bever, J. D., Westover, K. M. & Antonovics, J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J. Ecol.* **85**, 561–573 (1997).
- Van der Heijden, M. G. A. *et al.* Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* **396**, 69–72 (1998).
- Packer, A. & Clay, K. Soil pathogens and spatial patterns of seedling mortality in a temperate tree. *Nature* **404**, 278–281 (2000).
- Mills, K. E. & Bever, J. D. Maintenance of diversity within plant communities: soil pathogens as agents of negative feedback. *Ecology* **79**, 1595–1601 (1998).
- Callaway, R. M. & Ascheloug, E. T. Invasive plants versus their new and old neighbors: a mechanism for exotic invasion. *Science* **290**, 521–523 (2000).
- Olff, H., Hoorens, B., de Goede, R. G. M., van der Putten, W. H. & Gleichman, J. M. Small-scale shifting mosaics of two dominant grassland species: the possible role of soil-borne pathogens. *Oecologia* **125**, 45–54 (2000).
- Burdon, J. J. *Diseases and Plant Population Biology* (Cambridge Studies in Ecology, Cambridge Univ. Press, Cambridge, 1987).
- van der Putten, W. H. Pathogen-driven forest diversity. *Nature* **404**, 232–233 (2000).
- Blomqvist, M. M., Olff, H., Blaauw, M. B., Bongers, T. & van der Putten, W. H. Interactions between above- and belowground biota: importance for small-scale vegetation mosaics in a grassland ecosystem. *Oikos* **90**, 582–598 (2000).
- Klironomos, J. N., Moutoglou, P., Kendrick, B. & Widden, P. A comparison of spatial heterogeneity of vesicular-arbuscular mycorrhizal fungi in two maple-forest soils. *Can. J. Botany* **71**, 1472–1480 (1993).
- Bragulat, M. R., Abarca, M. L., Bruguera, M. T. & Cabanes, F. J. Dyes as fungal inhibitors: effect on colony diameter. *Appl. Environ. Microbiol.* **57**, 2777–2780 (1991).
- Domsch, K. H., Gams, W. & Anderson, T.-H. *Compendium of Soil Fungi* Vol. 1 (Academic, London, 1980).

Acknowledgements

I thank J. Bever, M. Hart, B. Husband and R. Reader for comments on the manuscript. I also thank A. Jones, G. LaPierre, G. Lewis, D. MacIntosh, J. Martin, J. Mosquin, P. Moutoglou and A. Shaw for laboratory and field assistance. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Competing interests statement

The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to J.N.K. (e-mail: jklirone@uoguelph.ca).

Effects of size and temperature on developmental time

James. F. Gillooly*, Eric L. Charnov*, Geoffrey B. West†‡, Van M. Savage†‡§ & James H. Brown*†

* Department of Biology, The University of New Mexico, Albuquerque, New Mexico 87131, USA

† Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA

‡ Theoretical Division, MS B285, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

§ Department of Physics, Washington University, St Louis, Missouri 63130, USA

Body size and temperature are the two most important variables affecting nearly all biological rates and times^{1–7}. The relationship of size and temperature to development is of particular interest, because during ontogeny size changes and temperature often varies^{8–12}. Here we derive a general model, based on first principles of allometry and biochemical kinetics, that predicts the time of ontogenetic development as a function of body mass and temperature. The model fits embryonic development times spanning a wide range of egg sizes and incubation temperatures for birds and aquatic ectotherms (fish, amphibians, aquatic insects and zooplankton). The model also describes nearly 75% of the variation in post-embryonic development among a diverse sample of zooplankton. The remaining variation is partially explained by stoichiometry, specifically the whole-body carbon to phosphorus ratio. Development in other animals at other life stages is also described by this model. These results suggest a general definition of biological time that is approximately invariant and common to all organisms.

The effects of body size and temperature on biological rates and times, including development time, have traditionally been studied separately. There is a rich literature on biological allometry that spans nearly a century^{1–3}. The relationships of various attributes of organisms such as metabolic rate, development time and lifespan, to body mass, m , are well approximated by power laws. In endothermic birds and mammals, where body temperature is nearly constant, biological rates and times (t) vary with body size as $t \propto m^{1/4}$ (refs 4 and 5). An equally rich literature on physiology relates many

Box 1

Relationship of equation (3) to Q_{10}

As most biological processes occur in the temperature range $T_c = 0–40^\circ\text{C}$, the term $(1 + (T_c/T_0))$ in equation (3) differs from unity by at most $40/273 \approx 0.15$. So equation (3) can be well approximated by:

$$a(T_c) = a(T_0)e^{(\bar{E}/kT_0^2)(T_c)}$$

Thus, mass-corrected development time ($t/m^{1/4}$, equation (5)) is inversely proportional to $\exp[(\bar{E}/kT_0^2)(T_c)]$ which becomes a Q_{10} when $T_c = 10^\circ\text{C}$. We then note that, because T_0 and k are fixed, Q_{10} depends only on $\bar{E}$, the activation energy. Taking an average value of $\bar{E} = 0.6 \text{ eV}$ gives $Q_{10} \approx e^{0.9} \approx 2.5$. This can also be expressed in terms of the slope of the fitted lines of Fig. 1a–d, which range from $\alpha = -(0.11–0.14)$ per $^\circ\text{C}$. Thus, Q_{10} can be approximated by $\exp(-10\alpha)$. Perhaps of greater importance is to recognize that the conventional Q_{10} factor is only an approximation. The exact expression includes further temperature dependence beyond the purely exponential dependence on T_c . Indeed, if Fig. 1a–d were replotted without the factor $(1 + T_c/T_0)$ in the exponent, the slopes would be 10–15% shallower, leading to a 10–15% error when using Q_{10} to obtain values of a .

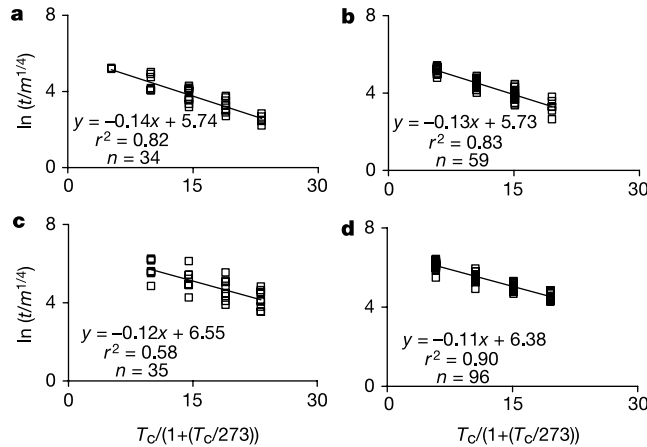


Figure 1 The effect of incubation temperature on mass-corrected embryonic development time for amphibians (a), fish (b), multivoltine aquatic insects (c) and zooplankton (d) incubated at different constant temperature. Incubation temperature is given as $T_c/[1 + (T_c/273)]$, in °C over the range 5–25 °C (see Methods); mass-corrected embryonic development time is given as $t/m^{1/4}$, in d per (mass at hatch in g) $^{1/4}$. Lines were fitted using least-squares linear regression. Data obtained from refs 8 and 9.

biological rates and times to temperature^{6,7}. Such temperature dependence is traditionally described in terms of Q_{10} , the assumed exponential change in rate for a temperature change of 10 °C. Recently, an allometric model for the effect of body size on growth was formulated based on the allocation of metabolic energy at the cellular level^{13,14}. The general equation is:

$$\frac{dm}{dt} = am^{3/4}[1 - (m/M)^{1/4}] \quad (1)$$

where the mass of the organism (m) as a function of time (t) is expressed in terms of the asymptotic mass (M) and a . The parameter a is related to fundamental cellular properties by $a = B_0 m_c/E_c$, where m_c is the mass of an average cell, E_c is the average amount of energy needed to create the cell, and B_0 is the normalization factor for metabolic rate¹³, B , which scales with mass as $B = B_0 m^{3/4}$. B_0 is proportional to the biochemical reaction rates for cellular metabolism, and therefore varies with temperature via a standard Boltzmann's factor $\exp(-\bar{E}/kT)$, where T is the absolute temperature (in K), $\bar{E}$ is the average energy for the reaction and k is Boltzmann's constant. As $a \propto B_0$, it has the same temperature dependence, namely $a(T) \propto \exp(-\bar{E}/kT)$. This can also be expressed in the conventional Q_{10} form (Box 1)¹⁵. The value of $a(T)$ at some temperature T is thereby related to its value at some other arbitrary

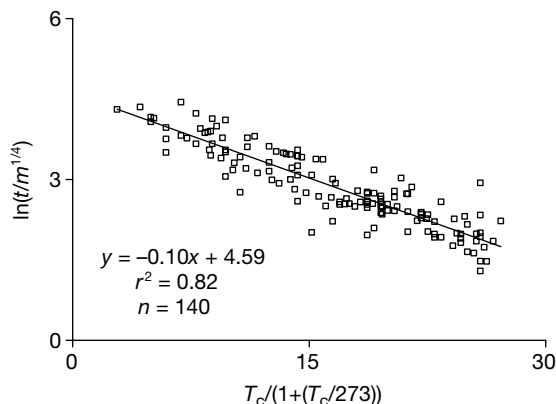


Figure 2 Plot as Fig. 1 but for marine fishes in the field (see Methods). Incubation temperatures ranged from 3 to 30 °C. The line is fit using least-squares linear regression. Data obtained from ref. 10.

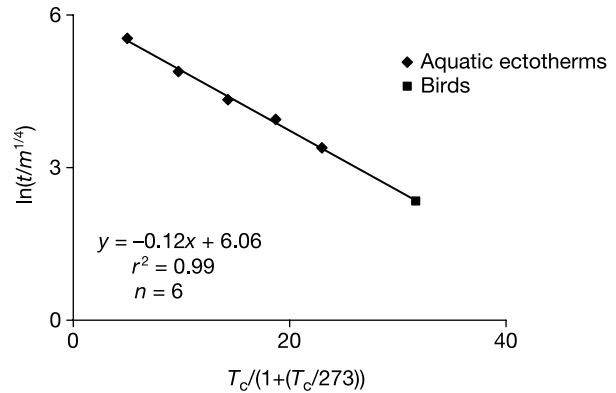


Figure 3 Plot as Fig. 1 but for aquatic ectotherms (data from Fig. 1a–d) and birds. The line is fit using least-squares linear regression to the mean values for all aquatic ectotherms (fish, amphibians, zooplankton, and aquatic insects; diamonds) and the mean value for birds (square) at different incubation temperatures ranging from 5 to 36 °C. Ectotherm data were obtained from refs 8 and 9, bird data from ref. 11.

temperature, T_0 , by $a(T)/a(T_0) = [\exp(-\bar{E}/kT)]/[\exp(-\bar{E}/kT_0)]$. Therefore,

$$a(T) = a(T_0)e^{-(\bar{E}/k)((1/T)-(1/T_0))} = a(T_0)e^{(\bar{E}/kT_0)((T-T_0)/T)} \quad (2)$$

Equation (2) can be expressed in terms of °C ($T_c = T - 273$) by setting $T_0 = 273$ K, the temperature at which water freezes and biological reactions cease. This yields the following:

$$a(T_c) = a(T_0)e^{(\bar{E}/kT_0^2)(T_c/(1+T_c/T_0))} \quad (3)$$

Throughout development, the mass of the embryo, m , is small compared to adult mass, M , so equation (1) is well approximated by $dm/dt = am^{3/4}$. When integrated from $m = 0$ at $t = 0$ at a fixed temperature this gives:

$$m = \left(\frac{a(T)t}{4}\right)^4 \quad \text{or} \quad \frac{t}{m^{1/4}} = \frac{4}{a(T)} \quad (4)$$

Substituting equation (3) into equation (4) gives:

$$\frac{t}{m^{1/4}} = \frac{4}{[a(T_0)e^{(\bar{E}/kT_0^2)(T_c/(1+T_c/T_0))}]} \quad (5)$$

which provides a general expression relating development time (t) to body mass (m) and temperature (T_c in °C). If body temperature T_c changes during growth, equation (1) must be integrated to reflect the time dependence of the parameter a .

Taking the logarithm of both sides of equation (5) predicts that plots of $\ln(t/m^{1/4})$ versus $T_c/(1+(T_c/273))$ will yield an approxi-

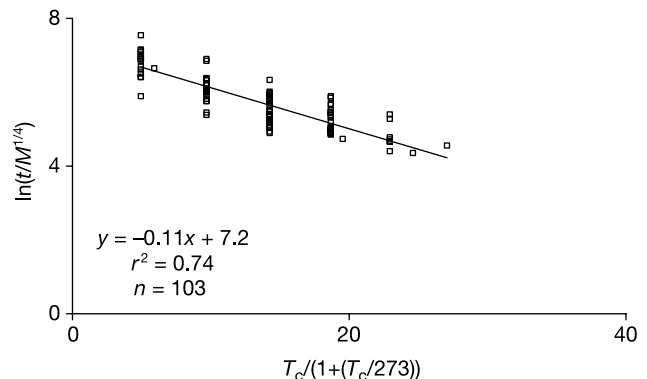


Figure 4 Plot as Fig. 1 but for post-embryonic (hatching to adult) development time for zooplankton (rotifers, copepods and cladocerans) incubated at different constant temperatures ranging from 5 to 30 °C. The line is fit using least-squares linear regression. Data sources listed in Methods.

mately universal straight line with slope, $\alpha = -\bar{E}/kT_0^2$, and intercept, $y_{\text{int}} = \ln[4/a(T_0)]$. As $\bar{E}$ and $a(T_0) = B_0 m_c/E_c$ depend on fundamental cellular properties, they, as well as the normalization factor B_0 , do not vary significantly across taxa¹⁵. Therefore, the slope (α) and intercept (y_{int}) should be approximately invariant quantities. We test this prediction using data from the simplest natural 'system' for growth: development time (zygote to hatchling) of eggs. Development of eggs is particularly well suited for assessing equation (5) because the embryo grows in mass as it incorporates the food stores in the egg. Furthermore, development occurs over a wide range of temperatures, from about 5 to 40 °C.

Using laboratory data on embryonic development time from four different groups of aquatic ectotherms (fish, amphibians, aquatic insects and zooplankton), plots of $\ln(t/m^{1/4})$ versus $T_c/(1 + (T_c/273))$ are indeed well fitted by straight lines with similar slopes and intercepts (Fig. 1a–d; see Methods)^{8,9}. Similar plots also fit field data on development time for marine fish eggs with a slope and intercept similar to the laboratory studies (Fig. 2) (see Methods)¹⁰. Furthermore, birds' eggs, which are incubated at much higher temperatures, are also fitted by such plots; pooled data for birds (13 orders, 172 species)¹¹, using a mean incubation temperature of 36 °C (ref. 8), fall on the same line as the pooled data for aquatic ectotherms (Fig. 3).

The success of the model in describing how size and temperature affect embryonic development time led us to consider whether it might also apply to development at different life stages. As shown in Box 2, the model makes similar predictions for post-embryonic development time (hatch to maturity). Nearly 75% of the variation

in post-embryonic development times in zooplankton can be explained by the model. This suggests that equation (5) applies to both embryonic and post-embryonic growth (Fig. 4) (see Methods)¹².

The regression lines for the mass-corrected relationships of development time to temperature (Figs 1–4) also provide an independent means of estimating the parameter a . The temperature dependence of a may be expressed in terms of the slope and intercept using equation (3): $a(T_c) = 4 \exp[-(\alpha T_c/(1 + T_c/T_0) + y_{\text{int}})]$. Taking average values of $\alpha = -0.12 \text{ per } ^\circ\text{C}$ and $y_{\text{int}} = 6 \ln(d \text{ g}^{-1/4})$ from Fig. 3, this equation predicts $a = 0.65 \text{ g}^{1/4} \text{ d}^{-1}$ for post-embryonic growth of birds at 40 °C, and $a = 0.018 \text{ g}^{1/4} \text{ d}^{-1}$ for post-embryonic growth of cod at 5 °C (ref. 16). These values compare favourably with independent estimates of a derived from fitting empirically measured growth curves¹⁴. These estimates give nearly the same value of $a = 0.017 \text{ g}^{1/4} \text{ d}^{-1}$ for the cod, and three values that bracket 0.65 for birds (0.47, 1.56, $1.90 \text{ g}^{1/4} \text{ d}^{-1}$)¹⁴. The fact that these two independent methods give similar values is further evidence that the model captures the effects of body size and temperature on growth.

Moreover, the activation energy for metabolic reactions can be used to predict the slope of the relationships in Figs 1–4. Using the equation $\alpha = -\bar{E}/kT_0^2$ (that is, equation (5)), and an average activation energy for metabolic reactions of 0.6 eV (range between approximately 0.2 and 1.2 eV)^{15,17–19}, we predict $\alpha = -0.09 \text{ per } ^\circ\text{C}$. The closeness of this value to the observed average value of $\alpha = -0.12 \text{ per } ^\circ\text{C}$ provides support for this model (that is, equation (5)).

We have therefore shown that body size and temperature account for much, but by no means all, of the variation in biological rates and times. Our model, based on first principles of allometry and kinetics, can help to isolate the causes of this still unexplained variation. For example, during post-embryonic growth, unlike embryonic growth, individuals must forage to obtain resources from environments where the availability of nutrients varies. In particular, it is suggested that organisms with higher mass-specific post-embryonic growth rates ($(1/m)(dm/dt)$) acquire more phosphorus (P) relative to other elements such as carbon (C) to produce the phosphorus-rich nucleic acids required for more frequent cell divisions and faster growth (that is, the 'stoichiometric growth hypothesis')^{20–22}. Thus, faster-growing organisms would be predicted to have lower C:P ratios. In Box 3, we relate this stoichiometric growth hypothesis to our size/temperature model. Figure 5 shows that C:P ratios explain much of the residual variation in Fig. 4.

Many biological times, including cardiac cycle, blood circulation time, and development time (that is, t), increase as the 1/4 power of

Box 2

Extension to post-embryonic growth

The general solution to equation (1) valid for all times is given by

$$\left(\frac{m}{M}\right)^{1/4} = 1 - \left[1 - \left(\frac{m_0}{M}\right)^{1/4}\right] e^{-at/4M^{1/4}}$$

where m_0 is the initial larval mass ($m = m_0$ at $t = 0$). Most zooplankton have determinate growth, and the onset of adulthood is assumed to be at $m = \delta M$; at this size growth ceases and the available energy is diverted to reproduction. We assume $\delta \approx 0.50$ – 0.90 , and is similar across species (see below). The time taken, t_m , to reach $m = \delta M$ is given by

$$\frac{t_m}{m^{1/4}} = \left(\frac{4}{a}\right) \left(\frac{1}{\delta^{1/4}}\right) \ln \left[\frac{(1 - (m_0/M)^{1/4})}{(1 - \delta^{1/4})} \right]$$

Apart from the $\delta^{1/4}$ term and the slowly varying logarithmic factor, this equation for post-embryonic growth (that is, hatch to maturity), is identical in structure to equation (5), which describes embryonic growth. Proceeding as before and using equation (3) for the temperature dependence of a implies that plots of $\ln(t_m/m^{1/4})$ versus $T_c/(1 + (T_c/273))$ will yield straight lines whose slopes are the same as those derived from equation (5) for embryonic growth: both should have slopes given by $\alpha = -\bar{E}/kT_0^2$. Their intercepts, on the other hand, should be slightly different: for post-embryonic growth the intercept is given by $\ln[4/a(T_0)] + \ln \ln[(1 - (m_0/M)^{1/4})/(1 - \delta^{1/4})] - 1/4 \ln \delta$ rather than simply $\ln[4/a(T_0)]$. The difference between these is rather small; if $m_0/M \ll 1$, δ in the range of 0.50–0.90 yields a correction in the range of approximately 0.50–1.3, a 10–22% increase in the intercept above the value of approximately 6 shown in Fig. 1. The correction depends very little on δ , so δ need not be strictly constant across species. Data for post-embryonic growth for a variety of zooplankton (rotifers, copepods and cladocerans) are plotted in this way in Fig. 4. As can be seen, a straight line is obtained with a slope of $-0.11 \text{ per } ^\circ\text{C}$ and an intercept of $7.2 \ln(\text{g}^{1/4} \text{ d}^{-1})$. Because the latter is 17% greater than 6, there is excellent agreement with corresponding values derived from embryonic growth data in Figs 1–3.

Box 3

The relationship to biological stoichiometry

To incorporate the 'stoichiometric growth hypothesis'^{20–22} we propose that $a(T)$ also depends on the C:P ratio so that $a(T) \propto \exp(-\bar{E}/kT) \lambda(\text{C} : \text{P})$, where $\lambda(\text{C} : \text{P})$ is a decreasing function of the C:P ratio. This can be used in equation (5) to predict that growth rates decrease with C:P across species (see Methods). To assess if C:P ratios do in fact decrease with growth rates across species, we plot C:P ratios against the corresponding residuals for zooplankton in Fig. 4 (Fig. 5). Species in Fig. 5 represent all major groups of zooplankton shown in Fig. 4 (cladocerans, calanoid and cyclopoid copepods), except rotifers for which stoichiometric data were not available. We predict an inverse relationship such that species that lie above the fitted line (that is, lower average growth rates) would have high C:P ratios, and vice versa. And, the plot does indeed show that considerable variation about the line in Fig. 4 is explained by differences in the C:P ratios among species. This supports our prediction and suggests that the relationship between size, temperature and biological stoichiometry proposed above may be correct.

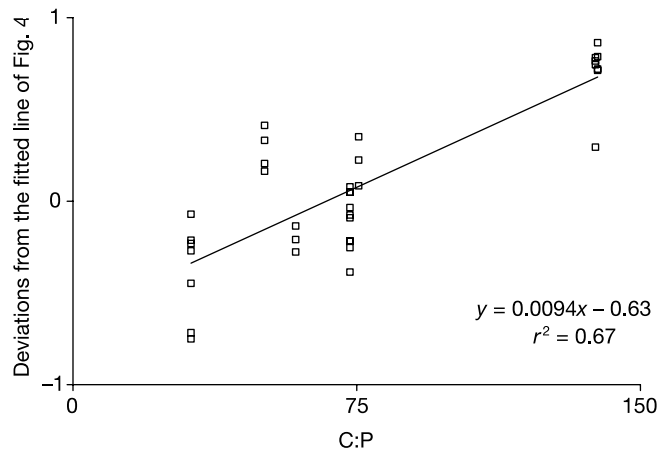


Figure 5 The relationship between deviations for the fitted line in Fig. 4 (that is, $T_c/(1 + (T_c/273))$ versus $t/m^{1/4}$) and whole-body carbon to phosphorus ratios (C:P) for adults of these species. Data sources listed in Methods.

body mass^{2,3}. Because all of these times are ultimately related to biochemical reaction rates, they are expected to decrease with temperature via the same Boltzmann factor, $\exp(-\tilde{E}/kT)$. Combined, the effects of mass and temperature therefore yield a general definition of biological time:

$$t_B = t(m/m_0)^{-1/4} e^{(-\tilde{E}/k(1/T - 1/T_0))} \\ = t(m/m_0)^{-1/4} e^{-\alpha T_c/(1 + T_c/T_0)} \quad (6)$$

where m_0 normalizes mass to some arbitrary value (for example, 1 g) and T_0 normalizes temperature to some arbitrary value (for example, 20 °C). This is the biological time clock. □

Methods

Embryonic development time

Embryonic development times of aquatic ectotherms were collected from compilations of laboratory studies where eggs were incubated at different constant temperatures ranging from 5 to 25 °C (refs 8 and 9). These include mostly freshwater, but some marine, species of both vertebrates and invertebrates (zooplankton: 2 phyla, 7 orders, 29 species; fishes: 7 orders, 21 species; amphibians: 2 orders, 10 species; multivoltine aquatic insects: 3 orders, 10 species). For each species, we included only data from the 'biologically relevant' temperature range required for normal development. Egg sizes were obtained from reference texts and used as an approximation for the mass of species at hatching, m , as the mass at hatch is not often measured. This introduces a maximum possible error of <5% so long as $m \geq 0.8$ times the egg mass. Methods are detailed in refs 8 and 9.

Field data on embryonic development times of marine fish are comparable to the laboratory data, except that T_c was taken as the 'prevailing temperature at incubation'¹⁰. Egg masses were calculated from egg diameters assuming a density of 1 g ml⁻¹ (ref. 8).

Post-embryonic development time and biological stoichiometry

Most post-embryonic development times and adult body masses were obtained from a compilation of published data¹², though some additional data were acquired for genera under-represented in this compilation. These include cladocerans (*Daphnia*^{23,24}, *Diaphanosoma*²⁴, *Ceriodaphnia*²⁵, *Bosmina*²⁶), and species of cyclopoid^{27,28} and calanoid copepods²⁹. Adult body masses for these species were estimated in the same manner as in the compilation. For Box 3, whole-body C:P ratios were obtained for as many of the species shown in Fig. 4 as possible. The stoichiometric ratios were published values for adults of those species, or adults of species from the same genus and similar body size^{21,30}.

Received 20 July 2001; accepted 31 January 2002.

- Schmidt-Nielsen, K. *Scaling: Why is Animal Size So Important?* (Cambridge Univ. Press, Cambridge, 1983).
- Calder, W. A. III *Size, Function, and Life History* (Harvard Univ. Press, Cambridge, Massachusetts, 1984).
- Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, Cambridge, 1983).
- Calder, W. A. III in *Avian Energetics* (ed. Paynter, R. A.) 86–151 (Nuttall Ornithology Club 15, Cambridge, 1974).
- Lindstedt, S. L. & Calder, W. A. III Body size, physiological time, and longevity of homeothermic animals. *Q. Rev. Biol.* **56**, 1–16 (1981).
- Somero, G. S. in *Handbook of Physiology* Vol. 13 (ed. Dantzler, W. H.) 1391–1444 (Oxford Univ. Press, New York, 1997).
- Cossins, A. H. & Bowler, K. *Temperature Biology of Animals* (Chapman and Hall, London, 1987).
- Gillooly, J. F. & Dodson, S. I. The relationship of neonate mass and incubation temperature to embryonic development time in a range of animal taxa. *J. Zool.* **251**, 369–375 (2000).
- Gillooly, J. F. & Dodson, S. I. The relationship of egg size and incubation temperature to embryonic

- development time in univoltine and multivoltine aquatic insects. *Freshwat. Biol.* **44**, 595–604 (2000).
- Pauly, D. & Pullin, R. S. V. Hatching time in spherical, pelagic, marine fish eggs in response to temperature and egg size. *Eviron. Biol. Fishes* **22**, 261–271 (1988).
- Heinroth, O. Die Beziehungen zwischen vogelgewicht, eigewicht, gelegewicht und brutdauer. *J. Ornithol.* **70**, 172–285 (1912).
- Gillooly, J. F. Effect of body size and temperature on generation time in zooplankton. *J. Plankton Res.* **22**, 241–251 (2000).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
- West, G. B., Brown, J. H. & Enquist, B. J. A general model for ontogenetic growth. *Nature* **413**, 628–631 (2001).
- Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M. & Charnov, E. L. Effects of size and temperature on metabolic rate. *Science* **293**, 2248–2251 (2001).
- Scott, W. B. & Scott, M. G. Atlantic fishes of Canada. *Can. Bull. Fish. Aquat. Sci.* **219** (1988).
- Vetter, R. A. H. Ecophysiological studies on citrate-synthase: I. Enzyme regulation of selected crustaceans with regard to temperature adaptation. *J. Comp. Physiol. B* **165**, 46–55 (1995).
- Raven, J. A. & Geider, R. J. Temperature and algal growth. *New Phytol.* **110**, 441–461 (1988).
- McLeese, J. M. & Eales, J. G. 3,5,3' Triiodo-L-thyroxine and L-thyroxine uptake into red blood cells of rainbow trout, *Oncorhynchus mykiss*. *Gen. Comp. Endocrinol.* **102**, 47–55 (1995).
- Elser, J. J., Dobberfuhl, D., MacKay, N. A. & Schampel, J. H. Organism size, life history and N:P stoichiometry: toward a unified view of cellular and ecosystem processes. *Bioscience* **46**, 674–684 (1996).
- Hessen, D. O. & Lyche, A. Inter- and intraspecific variations in zooplankton element composition. *Arch. Hydrobiol.* **121**, 343–353 (1991).
- Main, T. M., Dobberfuhl, D. R. & Elser, J. J. N:P stoichiometry and ontogeny of crustacean zooplankton: A test of the growth rate hypothesis. *Limnol. Oceanogr.* **42**, 1474–1478 (1997).
- Foran, J. A. A comparison of the life-history features of a temperate and a subtropical *Daphnia* species. *Oikos* **46**, 185–193 (1986).
- Hanazato, T. & Yasuno, M. Effect of temperature in the laboratory studies on growth, egg development and first parturition of five species of cladocera. *Jpn. J. Limnol.* **46**, 185–191 (1985).
- Andersen, D. H. & Benke, A. C. Growth and reproduction of the cladoceran *Ceriodaphnia dubia* from a forested floodplain swamp. *Limnol. Oceanogr.* **39**, 1517–1527 (1994).
- Kankaala, P. & Wulff, F. Experimental studies on temperature-dependent embryonic and postembryonic developmental rates of *Bosmina longispina maritima* (Cladocera) in the Baltic. *Oikos* **36**, 137–146 (1981).
- Maier, G. The effect of temperature on the development, reproduction, and longevity of two common cyclopoid copepods, *Eucyclops serrulatus* (Fischer) and *Cyclops strennus* (Fischer). *Hydrobiologia* **203**, 165–175 (1990).
- Munro, I. G. The effect of temperature on the development of egg, naupliar and copepodite stages of two species of copepods, *Cyclops vicinus* (Uljanin) and *Eudiaptomus gracilis* (Sars). *Oecologia* **16**, 355–367 (1974).
- Geiling, W. T. & Campbell, R. S. The effect of temperature on the development rate of the major life stages of *Diaptomus pallidus* (Herrick). *Limnol. Oceanogr.* **17**, 304–307 (1972).
- Andersen, T. & Hessen, D. O. Carbon, nitrogen, and phosphorus content of freshwater zooplankton. *Limnol. Oceanogr.* **36**, 807–814 (1991).

Acknowledgements

We thank S. Dodson, M. Ernest, C. M. Del Rio, E. Toolson, T. Turner and B. Wolf for comments or discussions that improved this manuscript. J.F.G. thanks J. S. Gillooly for support and encouragement. J.F.G., G.B.W. and J.H.B. are grateful for the support of the Thaw Charitable Trust and a Packard Interdisciplinary Science Grant; V.M.S., G.B.W. and J.H.B. for the support of the National Science Foundation; and E.L.C. for the support received as a MacArthur Fellow. G.B.W. also thanks the Theoretical Physics Department at Oxford University for its hospitality, and the EPSRC for support.

Correspondence and requests for materials should be addressed to J.F.G. (e-mail: gillooly@unm.edu).

Minocycline inhibits cytochrome *c* release and delays progression of amyotrophic lateral sclerosis in mice

Shan Zhu*, Irina G. Stavrovskaya†, Martin Drozda*, Betty Y. S. Kim*, Victor Ona*, Mingwei Li*, Satinder Sarang‡, Allen S. Liu*, Dean M. Hartley§, Du Chu Wu||, Steven Gullans‡, Robert J. Ferrante¶#, Serge Przedborski||☆, Bruce S. Kristal†** & Robert M. Friedlander*

* Neuroapoptosis Laboratory, Department of Neurosurgery, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

† Burke Medical Research Institute, White Plains, New York 10605, USA

‡ Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

§ Center for Neurologic Diseases, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

|| Department of Neurology, ☆ Department of Pathology, Columbia University, New York, New York 10032, USA

¶ Geriatric Research Education and Clinical Center, Bedford VA Medical Center, Bedford, Massachusetts 01730, USA

Neurology, Pathology, and Psychiatry Departments, Boston University School of Medicine, Boston, Massachusetts 02118, USA

** Departments of Biochemistry and Neuroscience, Weill Medical College of Cornell University, New York, New York 10021, USA

Minocycline mediates neuroprotection in experimental models of neurodegeneration. It inhibits the activity^{1–6} of caspase-1, caspase-3, inducible form of nitric oxide synthetase (iNOS) and p38 mitogen-activated protein kinase (MAPK). Although minocycline does not directly inhibit these enzymes, the effects may result from interference with upstream mechanisms resulting in their secondary activation. Because the above-mentioned factors are important in amyotrophic lateral sclerosis (ALS), we tested minocycline in mice with ALS^{7–9}. Here we report that minocycline delays disease onset and extends survival in ALS mice. Given the broad efficacy of minocycline, understanding its mechanisms of action is of great importance. We find that minocycline inhibits mitochondrial permeability-transition-mediated cytochrome *c* release. Minocycline-mediated inhibition of cytochrome *c* release is demonstrated *in vivo*, in cells, and in isolated mitochondria. Understanding the mechanism of action of minocycline will assist in the development and testing of more powerful and effective analogues. Because of the safety record of minocycline, and its ability to penetrate the blood–brain barrier, this drug may be a novel therapy for ALS¹⁰.

Minocycline is a second-generation tetracycline that effectively crosses the blood–brain barrier¹⁰. It has remarkable neuroprotective qualities in models of cerebral ischaemia, traumatic brain injury, and Huntington's and Parkinson's disease^{1–3,5,6}. Members of the caspase cell death family, in particular caspase-1 and caspase-3, are activated and have important roles in these diseases^{1,2,7,11}. In addition to caspases, iNOS is also upregulated and activated in these diseases^{1,2,8}. Minocycline-mediated neuroprotection is associated with inhibition of caspase-1, caspase-3 and iNOS transcriptional upregulation and activation^{1–3,5,6}. Inhibition of p38 MAPK and microglial activation have been associated with minocycline-mediated neuroprotection^{1,4}. Despite all these neuroprotective properties of minocycline, its precise primary target is not known. The goals of the present study are to evaluate the effectiveness of minocycline on a mouse model of ALS, and investigate its mechanism of neuroprotection.

ALS is a chronic neurodegenerative disease characterized by progressive motor weakness resulting from selective motor neuron cell death¹². Mortality is seen on the average four years following onset. The only proven therapy in humans, Riluzole, extends

survival in humans by approximately three months¹². It is therefore critical to identify new therapeutic strategies for ALS. Mutations in superoxide dismutase-1 (SOD1) have been linked to about 20% of patients with familial ALS¹³. Familial and sporadic forms of the disease have indistinguishable clinical and histopathological features. Transgenic mice expressing several of the mutant SOD1 genes found in humans with ALS develop motor neuron symptoms and histopathology resembling features of the human disease^{14,15}. Caspase-1 and caspase-3 are activated in neurons of ALS mice⁹. As caspases have an important role in ALS, and minocycline interferes with caspase pathways, we evaluated its effect on one of the ALS mouse models expressing the mutant human SOD1^{G93A} transgene. Minocycline (10 mg per kg body weight per day) was injected beginning at five weeks of age. Mutant hSOD1^{G93A} transgenic mice were evaluated weekly on a Rotarod to follow disease progression. Onset of impaired motor performance in minocycline-treated mice was delayed to 109.0 ± 1.5 days of age as compared to 90.3 ± 2.2 days in saline-treated mice ($P < 0.001$) (Fig. 1a–c). Minocycline administration extended survival from 125.6 ± 3.4 days to 136.8 ± 1.2 days ($P < 0.01$) (Fig. 1b, d). Minocycline-mediated neuroprotection in hSOD1^{G93A} mice was independently confirmed (by S.P.) using a slightly modified protocol. In this trial, minocycline administration started at six weeks of age at a slightly higher dose (11 mg per kg body weight per day). Mortality of saline-treated mice was 126.3 ± 2.7 days as compared to 139.0 ± 2.1 days of the minocycline-treated mice ($n = 8$ per group, $P < 0.05$). These results demonstrate effectiveness of minocycline in this transgenic ALS model.

We then investigated the mechanism of minocycline-mediated neuroprotection. The known activities of minocycline include inhibition of caspase-1 and caspase-3 transcriptional upregulation and activation. However, minocycline does not directly inhibit caspase-1 or caspase-3 activity². A challenge in elucidating the primary neuroprotective target of minocycline *in vivo*, especially in chronic neurodegenerative diseases, is that many pathways become activated during the progression of the disease as a result of a secondary reactive process. To determine the primary target (or targets) of minocycline, we searched for *in vitro* models where cell death is inhibited by minocycline. As minocycline mediates remarkable neuroprotection in ischaemia and traumatic brain injury^{1,3}, we evaluated its effect on *N*-methyl-D-aspartate (NMDA)-mediated cell death in primary cerebocortical neurons. Minocycline inhibited NMDA-induced cell death (Fig. 2a). In addition, minocycline inhibited death of SH-SY5Y neuroblastoma cells treated with either H₂O₂ or thapsigargin (THG) (Fig. 2b, c). Similar to its broad *in vivo* neuroprotective properties, minocycline inhibits cell death in a variety of *in vitro* models of cell death, suggesting that minocycline inhibits a critical shared step in the execution of the death program. As minocycline does not directly inhibit caspase-1 or caspase-3, we evaluated its effect on caspase-9 activation and cytochrome *c* release². Minocycline-mediated inhibition of THG-induced cell death correlates with inhibition of caspase-9 and of caspase-3 activation as well as of cytochrome *c* release (Fig. 2d, e).

Cytochrome *c* release from the mitochondria into the cytoplasm is a potent physiological stimulus for caspase-9 and caspase-3 activation¹⁶. Therefore, inhibition of cytochrome *c* release might be a primary target of minocycline. To directly probe this question, we evaluated the effect of minocycline on release of cytochrome *c* from cell-free mitochondrial preparations. Minocycline inhibited both calcium- and Bid-induced cytochrome *c* release in purified mouse liver mitochondria (Fig. 3a)^{17,18}. Owing to the critical role of mitochondria in apoptotic pathways, the relation between minocycline and inhibition of cytochrome *c* release is probably an important clue as to one of its primary/direct mechanisms of action¹⁶.

To confirm that mitochondria are direct targets of minocycline in the brain, and further probe into its mechanism of action, we evaluated the effect of minocycline on isolated non-synaptosomal

brain mitochondria. Cytochrome *c* release, changes in mitochondrial membrane potential ($\Delta\Psi$), Ca^{2+} transport, oxygen consumption and transmittance at 660 nm were monitored following Ca^{2+} addition (Fig. 3c–e and data not shown). As demonstrated in isolated liver mitochondria, and in the above-described models, minocycline inhibited cytochrome *c* release in isolated brain mitochondria. Minocycline addition was associated with alteration in several mitochondrial parameters, such as a decrease in $\Delta\Psi$ (Fig. 3d), but the change in transmittance (Fig. 3e) was the only change whose dose-specific effect reflected that of inhibition of cytochrome *c* release. Specifically, minocycline prevented most of the increase in transmittance that begins immediately on Ca^{2+}

addition. These studies, conducted in concentrated mitochondrial solutions appropriate for biochemical studies, required 200 μM minocycline for complete protection (Fig. 3a–e). But expressing minocycline levels as concentrations overstates requirements if minocycline is accumulated or acts stoichiometrically, in which case minocycline levels are more appropriately expressed in mol per

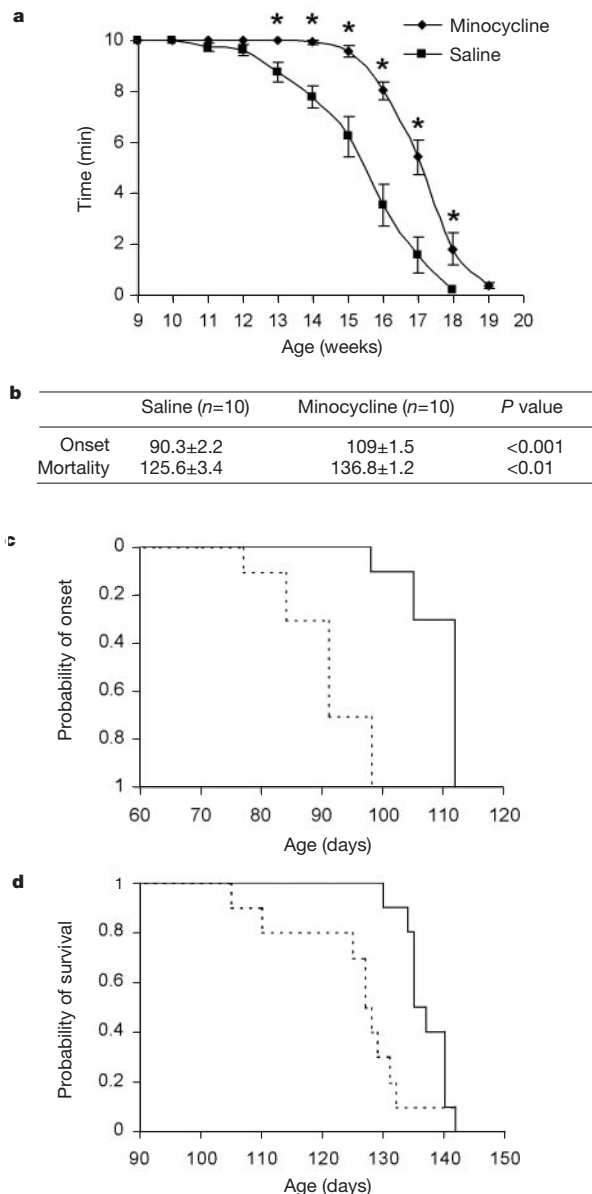


Figure 1 Minocycline delays onset and extends survival in ALS mice. **a**, Rotarod evaluation (see Methods) of *hSOD1*^{G93A} transgenic ALS mice treated with saline or minocycline ($n = 10$ per group; asterisk, $*P < 0.05$; error bars represent s.e.m.). **b**, Onset of motor deficits and mortality of ALS mice treated with minocycline or saline. **c**, **d**, Cumulative probability of onset of Rotarod deficits (**c**) and survival (**d**) in *hSOD1*^{G93A} mice. Onset of Rotarod deficit and mortality were significantly delayed in ALS mice treated with minocycline compared to mice treated with saline. Solid line, minocycline-treated ALS mice; dashed line, saline-treated control ALS mice.

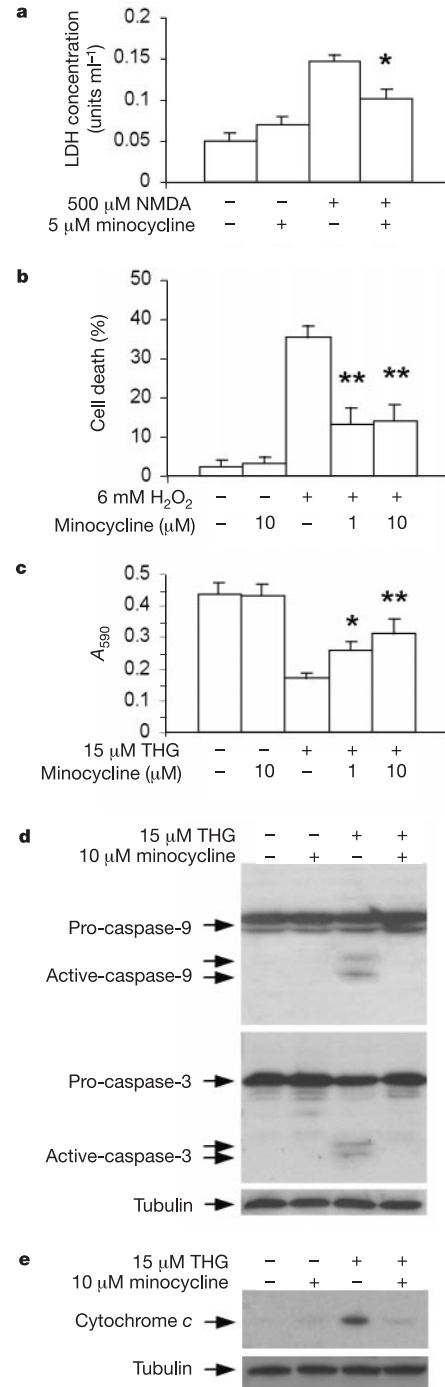


Figure 2 Minocycline inhibits cell death, caspase activation and cytochrome *c* release. **a**, Rat primary cortical neurons exposed to *N*-methyl-D-aspartate (NMDA) in the presence or absence of minocycline. **b**, **c**, H_2O_2 -mediated (**b**) and thapsigargin (THG)-mediated (**c**) SH-SY5Y cell death and its inhibition by minocycline. **d**, Caspase-9 and caspase-3 western blots of lysates of SH-SY5Y cells exposed to THG. **e**, Cytochrome *c* western blots of cytosolic fraction of SH-SY5Y cells exposed to THG. In cell death assays, data are representative of three independent experiments. (asterisk, $P < 0.05$; double asterisk, $P < 0.01$; error bars represent s.e.m.) A_{590} , absorbance at 590 nm.

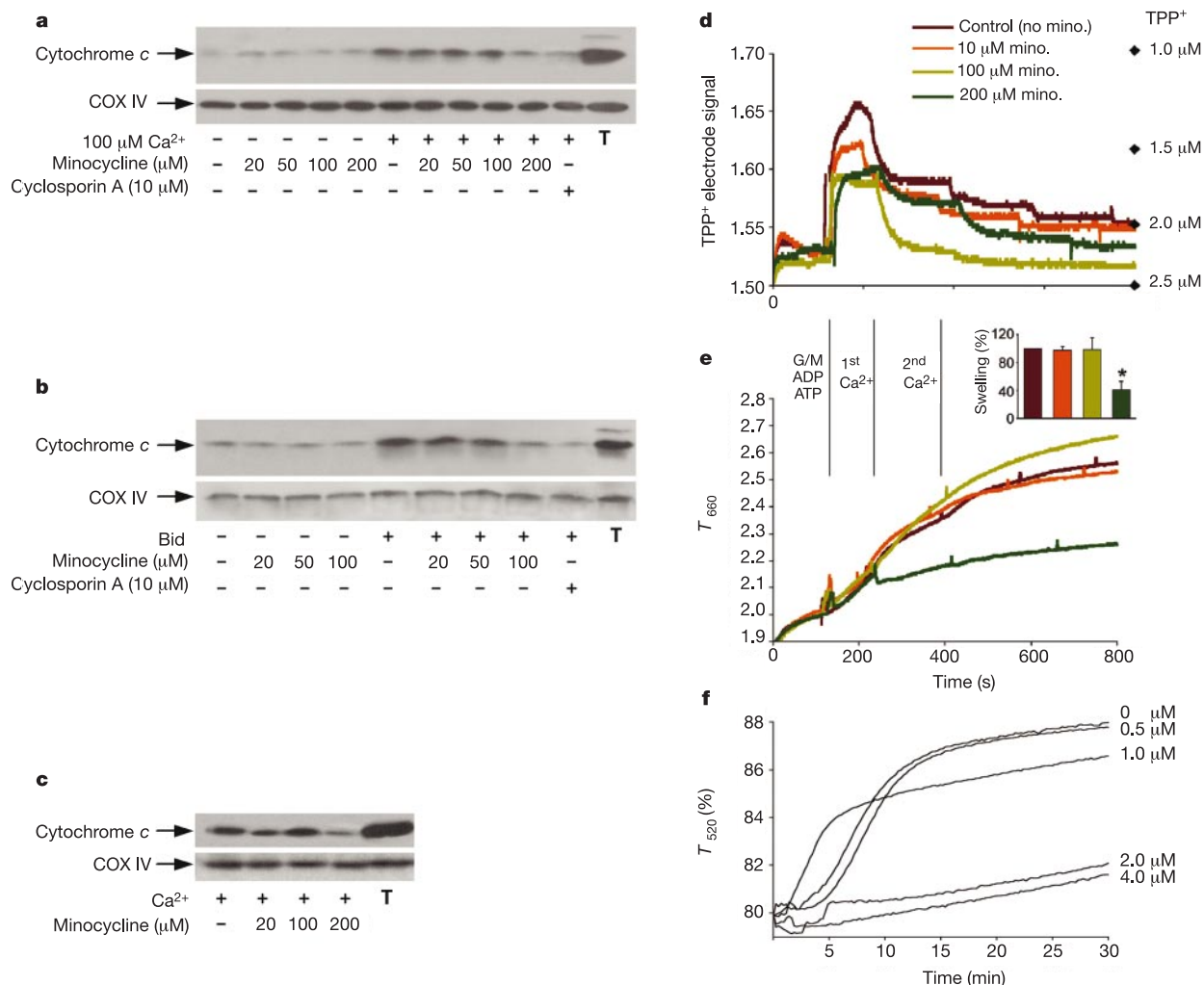


Figure 3 Minocycline inhibits cytochrome *c* release and swelling in purified mitochondria. **a,b**, Mouse liver mitochondria (0.5 mg ml⁻¹) were incubated with minocycline or cyclosporin A and cytochrome *c* release was induced by CaCl₂ (**a**) or purified Bid protein (**b**). T, total mitochondrial cytochrome *c*. COX IV was used as a loading control. **c–e**, Cytochrome *c* release (**c**) $\Delta\Psi$ (**d**) and light transmittance (**e**) in rat non-synaptosomal brain mitochondria (1 mg ml⁻¹) exposed to 40 μM Ca²⁺ pulses in the presence of varying minocycline concentrations. Inset in **e**, mean $\pm$ s.d. of the swelling (increase in

transmittance) relative to control at 600 s (shown as 100%). Bars coloured as main key. Asterisk, $P < 0.01$ versus 0, 10 and 100 μM minocycline, ANOVA followed by the statistically conservative Tukey/Kramer *post hoc*. **f**, Representative traces of effects of minocycline on swelling under conditions associated with PT induction in diluted rat liver mitochondria (0.03 mg ml⁻¹). Initial transmittance (T%) normalized to start at a common point (mean adjustment $<1\%$ of initial value).

mg protein. Co-titration of minocycline and mitochondria demonstrated minocycline-mediated inhibition of mitochondrial swelling at physiologically relevant concentrations of 2–4 μM in isolated rat liver mitochondria (≥ 40 nmol per mg protein, Fig. 3f, and data not shown). Traces are representative of a total of 24 experiments (3 independent mitochondrial preparations, 2 mitochondrial concentrations (0.1 mg ml⁻¹, 0.03 mg ml⁻¹), 2 substrates (glutamate/malate, succinate), 2 duplicates) on minocycline-mediated protection. For 0.1 mg per ml protein, 12 out of 12 samples were protected by 4 μM minocycline (40 nmol per mg protein); for 0.03 mg per ml protein, 12 out of 12 samples were protected by 2 μM minocycline (66 nmol per mg protein). The described minocycline-mediated effects were highly reproducible and statistically significant ($P < 0.0001$ analysis of variance, ANOVA, followed by the Fisher protected least significant difference (PLSD) *post hoc* test).

To confirm the findings of minocycline-mediated inhibition of cytochrome *c* release *in vivo*, we evaluated whether minocycline-mediated neuroprotection in ALS mice is associated with inhibition of cytochrome *c* release. Consistent with a previous report, release of cytochrome *c* in spinal cords of ALS mice correlates with disease

progression¹⁵. Minocycline treatment significantly inhibited release of cytochrome *c* in ALS mice ($n = 5$, $P < 0.05$) (Fig. 4a, b). As a confirmatory marker of efficacy, we show that minocycline inhibits caspase-3 activation in spinal cords of ALS mice (Fig. 4c). To extend these findings, we demonstrate that minocycline inhibits cerebral ischaemia-mediated cytochrome *c* release ($n = 5$, $P < 0.01$) (Fig. 4d, e). These results provide *in vivo* confirmation and support to the finding of minocycline-mediated inhibition of cytochrome *c* release in cells and in isolated mitochondria.

The data obtained are most consistent with a model in which minocycline acts directly on the mitochondria to alter PT-mediated cytochrome *c* release. Although other possibilities cannot be ruled out, several lines of evidence point to involvement of PT in the effects demonstrated by minocycline. Release of cytochrome *c* from isolated liver mitochondria occurs following exposure to exogenous Ca²⁺, the standard PT challenge, or Bid, which has been shown to accelerate PT induction, and is prevented in both cases by cyclosporin, the standard PT inhibitor^{19,20}. In non-synaptosomal mitochondria, minocycline failed to significantly block Ca²⁺-cycling driven loss of $\Delta\Psi$ (Fig. 3d and data not shown), but did consistently

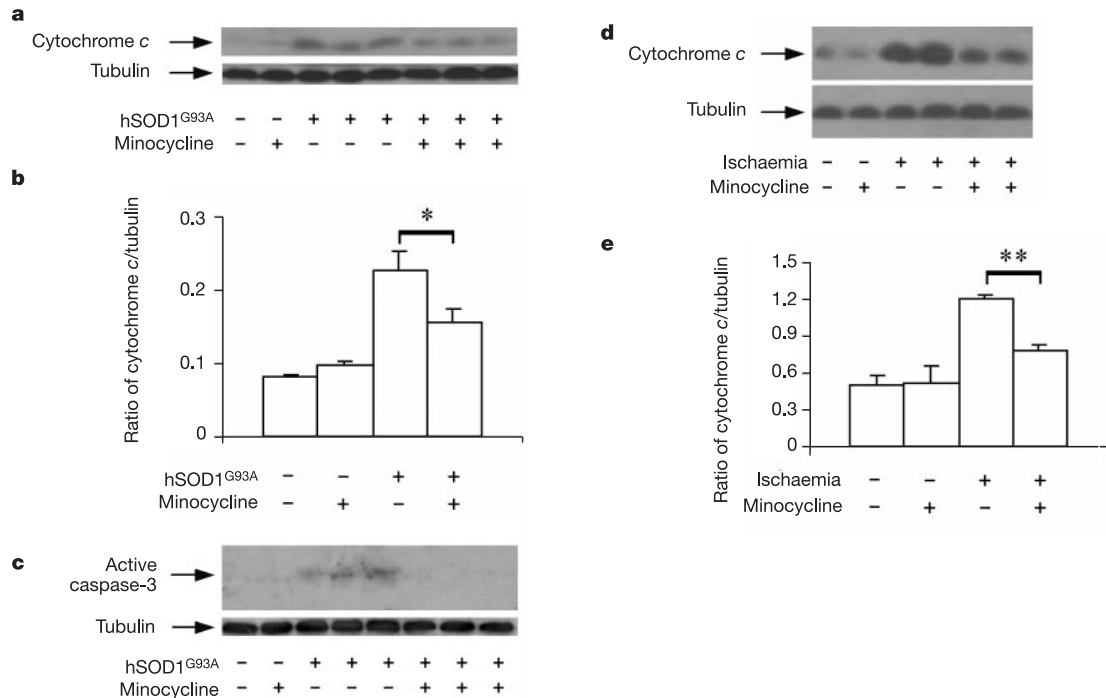


Figure 4 Minocycline inhibits cytochrome *c* release in ALS and ischaemia. **a,b**, ALS mice were injected with minocycline for 10 days starting at eight weeks of age. Spinal cord cytosolic components were evaluated by western blot for cytochrome *c* release (**a**). Densitometric quantification of cytochrome *c* release ($n = 5$, asterisk, $P < 0.05$) (**b**). **c**, Caspase-3 western blot of spinal cord lysate of ALS mice treated with/without

minocycline using an antibody specific for activated caspase-3 (p17). **d**, Cytosolic components of the ischaemic territory of mouse brains were fractionated and analysed for cytochrome *c* release. **e**, Densitometric quantification of cytochrome *c* release ($n = 5$, double asterisk, $P < 0.01$).

prevent over 50% of the total change in transmittance (Fig. 3e). Transmittance is the physical equivalent of the absorbance change considered a standard hallmark of PT induction in liver mitochondria and more controversially, one measure of 'PT-like' behaviour in non-synaptosomal preparations^{21,22}. Given that the exact mechanism mediating cytochrome *c* release remains unclear and controversial, these results do not rule out the possibility that alternative PT-independent pathways might exist, resulting in cytochrome *c* release^{16,20,23}.

We have validated the mechanism of action of minocycline at three levels: using cell-free mitochondrial preparations, using cell based death models, and *in vivo* using ALS mice and cerebral ischaemia models. As release of cytochrome *c* is a shared feature of many neurological disorders, inhibition of PT-mediated cytochrome *c* release by minocycline explains, at least in part, the broad inhibition of cell death (both *in vivo* and *in vitro*) mediated by this drug. Minocycline is, to our knowledge, the first non-toxic drug with a proven human safety record shown to inhibit cytochrome *c* release. Compounds such as minocycline will probably provide important synergism with compounds targeting other pathological mechanisms of disease progression^{2,9,24}. As demonstrated in this study, minocycline delays disease progression in a mouse model of ALS. Given its safety in chronic diseases, its oral bioavailability and its ability to cross the blood-brain barrier, minocycline could be evaluated for its effectiveness in human ALS¹⁰. Minocycline is at present being evaluated in human trials for Huntington's disease. Understanding the mechanism of action of minocycline will assist in the development and testing of more powerful and effective analogues. □

Methods

Minocycline treatment of ALS mice

ALS mice (Jackson Laboratories) were injected intraperitoneally daily with saline or

minocycline (Sigma). Strength and coordination were evaluated weekly by Rotarod (Columbus Instruments). Disease onset was defined as the first day a mouse could not remain on the Rotarod for 10 min at 15 r.p.m. Mortality was scored as age of death or age when the mouse was unable to right itself within 30 s (ref. 9).

NMDA neurotoxicity

Primary cortical neurons isolated from E16 rats and cultured as described²⁵. After two weeks, cultures were exposed for six hours to 500 μ M NMDA (Sigma) with or without a two-hour minocycline preincubation. Cytotoxicity was assayed by measuring lactate dehydrogenase (LDH) release²⁵.

Minocycline inhibition of SH-SY5Y cell death

Human neuroblastoma SH-SY5Y cells were preincubated with media containing minocycline for 24 h at 37 °C, and later exposed to 6 mM H₂O₂ for 4 h at 37 °C. Cells were then incubated with calcein-AM (1 μ M, Molecular Probes) in PBS for 40 min at 37 °C. Cell viability was determined using an LJI 96, 384 Analyst (Molecular Devices) fluorescence reader. Viability is converted to cell death ratio (Fig. 2b). For THG experiments, cells were preincubated with minocycline for 1 h and then exposed to 15 μ M THG (Sigma). After 12 h, cell death was evaluated by MTT assay (Roche).

Middle cerebral artery occlusion

C57/B6 mice (18–20 g) were injected with minocycline at 45 mg per kg body weight 4 h before ischaemia, then at 22.5 mg per kg body weight twice a day¹. After 120 min of middle cerebral artery occlusion, blood flow was restored²⁶. Brains were removed after 24 h and the ischaemic territory dissected for evaluation of cytochrome *c* release.

Tissue and cell cytosolic fractionation

Brains or spinal cord samples were homogenized (10 mM HEPES, pH 7.4, 250 mM sucrose, 10 mM KCl, 1.5 mM MgCl₂, 1 mM EDTA, 1 mM EGTA, 1 mM dithiothreitol (DTT), 50 μ M *N*-benzoxycarbonyl-Val-Ala-Asp-fluoromethylketone (zVAD-fmk)) plus protease inhibitor cocktail (Roche) in a Kontes dounce homogenizer with a B pestle (Kontes Glass Co.). Cytosolic component was fractionated as described²⁴. SH-SY5Y cell cytosolic fraction was prepared 6 h after THG treatment using the same protocol as for mouse brains.

Western blot

For caspase activation, samples were lysed in RIPA buffer with protease inhibitors. For cytochrome *c* release, cytosolic component from tissue (40 μ g) or cells (10 μ g) was loaded for evaluation. Cytochrome *c*, caspase-9 and caspase-3 antibodies were purchased from PharMingen; COX IV antibody from Clontech.

Liver and brain mitochondria preparation

Mouse liver mitochondria were prepared as described and resuspended in MRM buffer (250 mM sucrose, 10 mM HEPES, pH 7.5, 1 mM ATP, 5 mM sodium succinate, 80 μ M ADP, 2 mM K_2HPO_4) at a concentration of 0.5 mg ml⁻¹ (ref. 18). Rat liver mitochondria were isolated from 4–6-month-old Fischer 344 $\times$ brown Norway F₁ rats by differential centrifugation as described, except the final wash and resuspension buffer had no EGTA, EDTA or BSA²⁷. Non-synaptosomal rat brain mitochondria were prepared from forebrains of ~8-week Fischer 344 $\times$ brown Norway rats (Ficoll gradient purification)^{28,29}.

In vitro cytochrome c release

An aliquot of 25 μ l of 0.5 mg ml⁻¹ mouse liver mitochondrial preparation was preincubated with minocycline or cyclosporin A for 5 min in MRM buffer. Mitochondria were incubated with 100 μ M CaCl₂ or 100 ng of purified mouse Bid protein¹⁷ at 30 °C for 30 min (CaCl₂) or for one hour (Bid). Mixtures were centrifuged at 10,000g at 4 °C for 10 min and the supernatant evaluated by western blot.

$\Delta\Psi$ /swelling measurements

Rat brain mitochondria (1 mg ml⁻¹ in 100 mM KCl, 75 mM mannitol, 25 mM sucrose, 10 mM Tris, 1 mM K-PO₄ (pH 7.3)) containing tetraphenyl phosphonium (TPP⁺), 25 μ M Ca²⁺ (from buffer and TPP⁺ stock) and minocycline (see key in Fig. 3d, e) were added at $t = 0$ and energized where marked (Fig. 3d, e) with 5 mM glutamate, 5 mM malate, 1 mM ATP and 80 μ M ADP. Bolus doses of 40 μ M Ca²⁺ were added where shown (Fig. 3d, e). Simultaneous measurement of $\Delta\Psi$ and light transmittance in stirred samples was accomplished using a four-channel respiration system designed by B. Krasnikov (ref. 30, and manuscript in preparation). Oxygen uptake, $\Delta\Psi$, and Ca²⁺ were measured using Clark, TPP⁺, and Ca²⁺-sensitive electrodes, respectively. A₆₆₀ was measured using a diode. Experiments were carried out in triplicate. Plots shown are representative. Mixtures were sampled and centrifuged at 10,000g at 4 °C for 10 min. Supernatants were evaluated for cytochrome c (Fig. 3c).

Rat liver mitochondria were used for minocycline-mitochondria co-titration. PT induction was assessed spectrophotometrically by suspending mitochondria at 25 °C in 200 μ l of 310 mM sucrose, 30 mM KCl, 3 mM K-HEPES (pH 7.3), with 5 μ M added Ca²⁺. Samples with 0.1 mg mitochondria per ml had 100 μ M K-PO₄; samples with 0.03 mg mitochondria per ml had 30 μ M K-PO₄. Changes in transmittance at 520 nm were followed for 30 min using a SpectraMax 250 Plate Reader (Molecular Dynamics). Minocycline does not display appreciable absorbance at this wavelength.

Light scattering data are qualitatively identical at the two wavelengths used. The lower wavelength (520 nm) was used on the plate reader because it gives a slightly better signal:noise profile. The higher wavelength (660 nm) was used in the four-channel system because the light-emitting diode can only be used at that wavelength.

Received 21 February; accepted 5 April 2002.

- Yrjanheikki, J., Keinänen, R., Pellikka, M., Hokfelt, T. & Koistinaho, J. Tetracyclines inhibit microglial activation and are neuroprotective in global brain ischemia. *Proc. Natl Acad. Sci. USA* **95**, 15769–15774 (1998).
- Chen, M. *et al.* Minocycline inhibits caspase-1 and caspase-3 expression and delays mortality in a transgenic mouse model of Huntington disease. *Nature Med.* **6**, 797–801 (2000).
- Sanchez Mejia, R. O., Ona, V. O., Li, M. & Friedlander, R. M. Minocycline reduces traumatic brain injury-mediated caspase-1 activation, tissue damage, and neurological dysfunction. *Neurosurgery* **48**, 1393–1401 (2001).
- Tikka, T., Fiebich, B. L., Goldsteins, G., Keinänen, R. & Koistinaho, J. Minocycline, a tetracycline derivative, is neuroprotective against excitotoxicity by inhibiting activation and proliferation of microglia. *J. Neurosci.* **21**, 2580–2588 (2001).
- Wu, D. C. *et al.* Blockade of microglial activation is neuroprotective in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine mouse model of Parkinson disease. *J. Neurosci.* **22**, 1763–1771 (2002).
- Du, Y. *et al.* Minocycline prevents nigrostriatal dopaminergic neurodegeneration in the MPTP model of Parkinson's disease. *Proc. Natl Acad. Sci. USA* **98**, 14669–14674 (2001).
- Friedlander, R. M., Brown, R. H., Gagliardini, V., Wang, J. & Yuan, J. Inhibition of ICE slows ALS in mice. *Nature* **388**, 31 (1997).
- Almer, G., Vukosavic, S., Romero, N. & Przedborski, S. Inducible nitric oxide synthase up-regulation in a transgenic mouse model of familial amyotrophic lateral sclerosis. *J. Neurochem.* **72**, 2415–2425 (1999).
- Li, M. *et al.* Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model. *Science* **288**, 335–339 (2000).
- Brogden, R. N., Speight, T. M. & Avery, G. S. Minocycline: A review of its antibacterial and pharmacokinetic properties and therapeutic use. *Drugs* **9**, 251–291 (1975).
- Ona, V. O. *et al.* Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* **399**, 263–267 (1999).
- Rowland, L. P. & Schneider, N. A. Amyotrophic lateral sclerosis. *N. Engl. J. Med.* **344**, 1688–1700 (2001).
- Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362**, 59–62 (1993).
- Gurney, M. E. *et al.* Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science* **264**, 1772–1775 (1994).
- Wong, P. C. *et al.* An adverse property of a familial ALS-linked SOD1 mutation causes motor neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron* **14**, 1105–1116 (1995).
- Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
- Li, H., Zhu, H., Xu, C. J. & Yuan, J. Cleavage of BID by caspase 8 mediates the mitochondrial damage in the Fas pathway of apoptosis. *Cell* **94**, 491–501 (1998).
- Luo, X., Budihardjo, I., Zou, H., Slaughter, C. & Wang, X. Bid, a Bcl2 interacting protein, mediates cytochrome c release from mitochondria in response to activation of cell surface death receptors. *Cell* **94**, 481–490 (1998).
- Zamzami, N. *et al.* Bid acts on the permeability transition pore complex to induce apoptosis. *Oncogene* **19**, 6342–6350 (2000).

- Bernardi, P., Scorrano, L., Colonna, R., Petronilli, V. & Di Lisa, F. Mitochondria and cell death. Mechanistic aspects and methodological issues. *Eur. J. Biochem.* **264**, 687–701 (1999).
- Kristal, B. S. & Dubinsky, J. M. Mitochondrial permeability transition in the central nervous system: induction by calcium cycling-dependent and -independent pathways. *J. Neurochem.* **69**, 524–538 (1997).
- Friberg, H., Connern, C., Halestrap, A. P. & Wieloch, T. Differences in the activation of the mitochondrial permeability transition among brain regions in the rat correlate with selective vulnerability. *J. Neurochem.* **72**, 2488–2497 (1999).
- Shimizu, S., Narita, M. & Tsujimoto, Y. Bcl-2 family proteins regulate the release of apoptogenic cytochrome c by the mitochondrial channel VDAC. *Nature* **399**, 483–487 (1999).
- Guegan, C., Vila, M., Rosoklija, G., Hays, A. P. & Przedborski, S. Recruitment of the mitochondrial-dependent apoptotic pathway in amyotrophic lateral sclerosis. *J. Neurosci.* **21**, 6569–6576 (2001).
- Hartley, D. M. *et al.* Protofibrillar intermediates of amyloid beta-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons. *J. Neurosci.* **19**, 8876–8884 (1999).
- Friedlander, R. M. *et al.* Expression of a dominant negative mutant of interleukin-1 beta converting enzyme in transgenic mice prevents neuronal cell death induced by trophic factor withdrawal and ischemic brain injury. *J. Exp. Med.* **185**, 933–940 (1997).
- Kristal, B. S. & Brown, A. M. Apoptogenic ganglioside GD3 directly induces the mitochondrial permeability transition. *J. Biol. Chem.* **274**, 23169–23175 (1999).
- Lai, J. C. & Clark, J. B. Preparation of synaptic and nonsynaptic mitochondria from mammalian brain. *Methods Enzymol.* **55**, 51–60 (1979).
- Kristal, B. S., Staats, P. N. & Shestopalov, A. I. Biochemical characterization of the mitochondrial permeability transition in isolated forebrain mitochondria. *Dev. Neurosci.* **22**, 376–383 (2000).
- Krasnikov, B. F., Kuzminova, A. E. & Zorov, D. B. The Ca²⁺-induced pore opening in mitochondria energized by succinate-ferricyanide electron transport. *FEBS Lett.* **419**, 137–140 (1997).

Acknowledgements

We thank E. Friedlander for editorial assistance, B. Krasnikov for discussion concerning the 4-channel mitochondrial chamber, and M. Lukyanova for technical assistance. Mouse Bid expression construct was provided by H. Li and J. Yuan. This work was supported by Project A.L.S. (R.M.F., S.G., S.P.), the NIH (R.M.F., D.M.H., R.J.F., S.P., B.S.K.), the Huntington's Disease Society of America (R.M.F.), the Hereditary Disease Foundation (R.M.F., B.S.K.), the Muscular Dystrophy Association (R.M.F., S.P.), the Veterans Administration (R.J.F.), Hope for ALS (S.G.), Ride for ALS (S.G.), ALS Association (S.P.), the US Department of Defense (S.P.), the Lowenstein Foundation (S.P.), the Smart Foundation (S.P.), and the Parkinson's Disease Foundation (S.P.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M.F. (e-mail: rfriedlander@rics.bwh.harvard.edu).

Extensive and divergent circadian gene expression in liver and heart

Kai-Florian Storch^{*}, Ovidiu Lipan[†], Igor Leykin[†], N. Viswanathan[‡], Fred C. Davis[‡], Wing H. Wong^{†§} & Charles J. Weitz^{*}

^{*} Department of Neurobiology, Harvard Medical School; [†] Department of Biostatistics, Harvard School of Public Health; and [‡] Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA
[§] Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA

Many mammalian peripheral tissues have circadian clocks^{1–4}; endogenous oscillators that generate transcriptional rhythms thought to be important for the daily timing of physiological processes^{5,6}. The extent of circadian gene regulation in peripheral tissues is unclear, and to what degree circadian regulation in different tissues involves common or specialized pathways is unknown. Here we report a comparative analysis of circadian gene expression *in vivo* in mouse liver and heart using oligonucleotide arrays representing 12,488 genes. We find that peripheral circadian gene regulation is extensive (≥ 8 –10% of the genes expressed in each tissue), that the distributions of circadian phases in the two tissues are markedly different, and that very

few genes show circadian regulation in both tissues. This specificity of circadian regulation cannot be accounted for by tissue-specific gene expression. Despite this divergence, the clock-regulated genes in liver and heart participate in overlapping, extremely diverse processes. A core set of 37 genes with similar circadian regulation in both tissues includes candidates for new clock genes and output genes, and it contains genes responsive to circulating factors with circadian or diurnal rhythms.

Mice were entrained (synchronized) to a 12-h light/dark cycle for ≥ 2 weeks, and then placed in constant dim light (less than 1 lx) for ≥ 42 h. We collected tissues at 4-h intervals over two circadian cycles (12 time points). RNA isolated from livers and hearts was analysed on oligonucleotide arrays^{7,8} representing 30–40% of the estimated total number of mouse genes^{9,10}.

We identified circadian patterns in the gene expression data by a series of filtering steps. For each gene we required: (1) a positive correlation between the first and second days (selects for a daily repetition); (2) a negative autocorrelation at 12 h (removes ultradian oscillations and noise); (3) an amplitude threshold (removes trivial oscillations); and (4) a threshold for the maximum signal (removes unreliable data). To optimize the filtering parameters, we designated nine 'guide genes' (*Per1*, *Per2*, *Per3*, *Bmal1* (also known as *Mop3*), *Tef*, *Dpb*, *E4bp4*, *Cry1* and *Cry2*) that are known to exhibit circadian regulation in both tissues. The guide genes provided a variety of phases, amplitudes and waveforms. We systematically varied the parameters to find combinations that minimized the overall number of genes selected but maximized the number of guide genes selected. We identified parameters that retained only about 4% of the genes on the array but $\geq 67\%$ of the guide genes. Further increasing the stringency resulted in sharp reductions in the selection of guide genes and other genes.

This procedure identified 575 genes in liver and 462 genes in heart with circadian expression patterns (Fig. 1). The oscillations showed a variety of phases, amplitudes and waveforms. To estimate the contribution of noise, we created control data sets identical in content to the experimental data set, by random permutation of time order and redistribution of amplitudes and means. Filtering 100 different control data sets for each tissue with the same parameters as above resulted in selection of 90 ± 9.2 (s.d.) genes for liver and 54 ± 7.8 genes for heart. Thus we estimate that 16% of our selected genes for liver and 12% for heart can be ascribed to noise. Of 51 genes regulated by a circadian clock recently identified in mouse liver by differential display¹¹, 15 were represented on the array, and none were guide genes. Of these, 8 were included in the liver circadian set. It seems likely that our analysis has identified a substantial fraction of clock-regulated genes but probably provides an underestimate.

Based on a probability of detection (see http://www.affymetrix.com/products/algorithms_tech.html for statistical algorithms for monitoring gene expression on GeneChip probe arrays) of $P < 0.05$ on at least 7 of the 12 arrays, we classified as expressed a total of 4,805 genes in liver and 5,120 in heart (a total of 3,953 genes in common). Given that 30–40% of all mouse genes are represented on the array, our estimates agree well with evidence that most cell types express 10,000–15,000 genes¹². After accounting for noise, we calculate that approximately 10% of the genes expressed in liver and about 8% of the genes expressed in heart show circadian regulation. Our data suggest a considerably greater prevalence of circadian regulation at low amplitudes, particularly in heart (data not shown). Although it is probable that the clock within each tissue drives most circadian gene regulation in the tissue, our *in vivo* analysis cannot distinguish genes driven locally from those driven by circadian signals generated in the suprachiasmatic nucleus¹³ or elsewhere. In a fibroblast cell line, a model for local circadian regulation, about 2% of the expressed genes were regulated by a circadian clock¹⁴, but this is probably a substantial underestimate¹⁴ because of damping of circadian oscillations¹.

The expression profiles of the circadian gene sets are displayed in Fig. 1. In liver (Fig. 1a), the oscillations showed a broad distribution of phases (Fig. 1b). In contrast, in heart (Fig. 1c) most of the genes showed synchronous peak expression at a circadian time of about 4 h (Fig. 1d, CT 4), a phase very sparsely represented in the liver set. This marked difference in circadian phase distribution was unaffected by significant relaxation or tightening of the filtering parameters. Thus, despite a similar number of genes with circadian expression in liver and heart, the timing of circadian gene regulation in the two tissues is markedly different, even though the molecular clocks in the two tissues were synchronized (Fig. 1b, d; phase markers).

The liver and heart circadian gene sets revealed very little overlap, with only 52 genes in common. Of these, a core set of only 37 showed essentially identical phases of peak expression in both

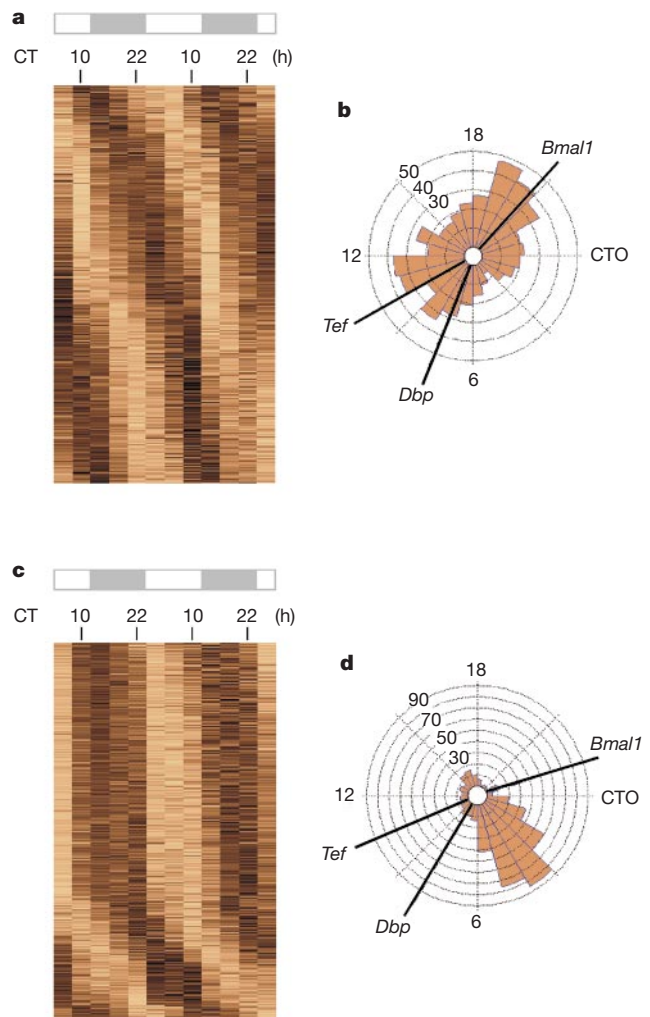


Figure 1 Temporal profiles of circadian gene expression in liver (**a, b**) and heart (**c, d**). **a, c**, Each column represents a circadian time point and each row a gene, with the genes ordered by the phase of maximal expression as determined by Fourier analysis. Light shades represent expression values above the mean for the gene; dark shades below the mean. Numbers at the top represent the circadian time (CT), and the bar represents the time of subjective day (white) and subjective night (grey). For a complete list of genes and associated expression values, see Supplementary Information. **b, d**, Phase histograms for genes with circadian expression patterns in liver and heart. Numbers on nested circles represent numbers of genes. CT0, circadian time zero hour. *Bmal1*, *Dpb* and *Tef* are genes with known robust circadian expression serving as phase markers for the circadian clock within each tissue.

tissues. This remarkable divergence in circadian gene regulation could simply reflect preferential regulation of tissue-specific genes by the clock in each tissue. We therefore examined the genes showing circadian expression only in liver or only in heart for detectable expression in the other. Of the 523 genes with circadian regulation only in liver, 316 (60%) were detectably expressed in heart. Conversely, of the 410 genes with circadian regulation only in heart, 242 (59%) were detectably expressed in liver. Thus circadian regulation of tissue-specific genes cannot account for the divergence of circadian gene regulation between liver and heart.

The divergence of liver and heart in the timing and content of circadian gene regulation suggests a specialized role of circadian clocks in each tissue, but the sheer extent of circadian gene regulation in both tissues implies a broad function. To address this apparent contradiction, we used the gene ontology representation¹⁵, which organizes biological knowledge into three hierarchies: biological process, molecular function and cellular component. Gene ontology provides a general and objective basis for analysing and comparing large biological data sets.

First we mapped the genes in our various sets onto the gene ontology hierarchies. Of the 4,805 genes expressed in liver and the 5,120 expressed in heart, 3,023 (63%) and 3,182 (62%), respectively, made a match to at least one gene ontology term; each term corresponds to a node (category) on a gene ontology hierarchy. Of the 575 genes in the liver circadian set and the 462 genes in the heart circadian set, 388 (67%) and 297 (64%), respectively, matched at least one gene ontology term, proportions similar to the expressed sets.

To create a framework for comparing the functions of liver and heart circadian gene regulation, we merged the set of genes expressed in liver with the set expressed in heart, and mapped the combined set onto the gene ontology hierarchy for biological process. The resulting map can be displayed as a tree that shows

only those nodes (biological process terms) represented by genes in the combined liver and heart expressed gene set. Figure 2 shows this tree at low resolution (grey structure with dots representing nodes). The most general processes are represented at the top (examples are marked by numbers); the most specialized processes are shown at the bottom. Onto this tree we superimposed the nodes making a match to at least one gene in the circadian set for liver (green dots) and those making a match to at least one gene in the circadian set for heart (red dots)—many make a match to both (yellow dots).

Several conclusions emerge from this global comparison. First, the circadian gene sets from both liver and heart map across the entire tree. Thus circadian clocks influence extremely diverse processes, even within a single tissue. Second, genes in the circadian sets for liver and heart map to many nodes in common, even at deep levels representing highly specialized processes. This result was robust across different filtering and mapping parameters. A few exceptions were evident, such as the cluster of nodes representing amino-acid metabolism (liver only; green arrow) and the cluster representing G-protein-coupled receptor signalling cascades (heart only; red arrow). We conclude that clock-regulated genes in liver and heart participate in many related or overlapping processes, even though the two sets of genes have almost no overlap in content or temporal expression pattern.

Our results, which address circadian gene regulation, provide a broad view of the probable physiological functions of peripheral clocks; however, some caveats apply. First, because protein rhythms depend in part on protein turnover, a transcript with a circadian rhythm does not necessarily code for a protein with a significant circadian rhythm; conversely, a transcript with an undetected low-amplitude circadian rhythm could code for a protein with a significant circadian rhythm. Second, a protein circadian rhythm might have little physiological consequence if, at its trough, the protein is not limiting or if it does not participate in a rate-limiting

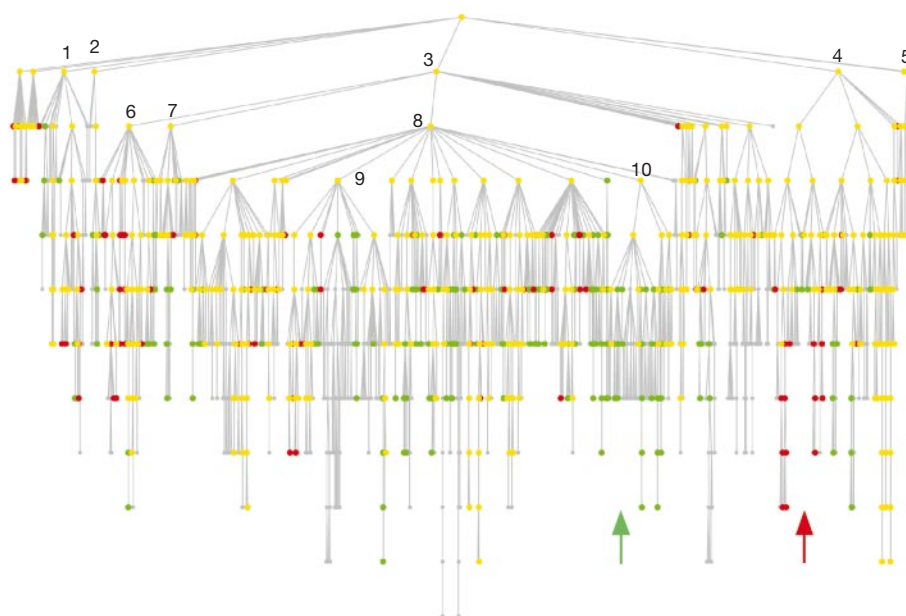


Figure 2 Global comparison of biological processes associated with the genes exhibiting circadian expression in liver and heart. The tree (grey dots and connecting paths) represents those categories (dots) in the gene ontology hierarchy for biological process that matched one or more genes expressed in liver or heart. Superimposed on the tree are the biological processes that matched one or more genes in the liver circadian set (green dots), the heart circadian set (red dots), or both (yellow dots). Marked are atypical examples of process clusters matching only to the circadian gene set from liver (green

arrow) or heart (red arrow). Selected examples of biological process categories are indicated by the following numbers: 1, developmental process; 2, death; 3, cell growth and/or maintenance; 4, cell communication; 5, behaviour; 6, transport; 7, stress response; 8, metabolism; 9, (metabolism) nucleobase/nucleoside/nucleotide; 10, (metabolism) amino-acid derivative. More than 1,400 biological process categories are shown.

process; conversely, a low-amplitude protein rhythm might drive a high-amplitude physiological rhythm if the protein participates in a cooperative process.

Because the set of genes with circadian regulation in both liver and heart is so small, it is likely to be highly enriched in genes important for the core functions of all peripheral clocks, such as synchronization to circulating factors, generation of rhythmicity, and control of transcriptional outputs. The expression profiles of the genes in the core common set are shown in Fig. 3a, arranged by the phase of peak expression as determined for liver. The slight disorder in the heart profiles arises from minor discrepancies between liver and heart in the phase assignments for some genes. The distribution of phases is essentially bimodal (Fig. 3b), with most genes showing peak expression between circadian time 6 h and 14 h, and a smaller group peaking in phase at about circadian time 20 h. Comparisons of temporal expression patterns for individual genes in liver and heart are shown in Fig. 4; in many cases the waveform is conserved in the two tissues.

Six of the nine guide genes appear in the common set (Fig. 4), but *Per3*, *Cry1* and *Cry2* do not because our filtering parameters excluded them from one or both circadian sets on the basis of low expression and/or amplitude. Also appearing (but not a guide gene) was the nuclear receptor gene *Rev-erb-α* (Fig. 4), which is known to be clock-regulated in different tissues¹⁶.

The other genes in the core common set were not previously known to exhibit circadian expression. *Zfp36* and *Sox3*, coding for transcription factors, are potential clock genes or core output genes, similar to *Dbp*¹⁷. Genes involved in chromatin regulation (*Chrac1*, *Rad23b* homologue, DNMT1-associated protein-1) and in ubiquitin pathways (*Usp2*, Fig. 4, and *Herpud1*) suggest a general circadian influence on gene expression and protein stability, respectively. The signal-transducing kinases *Gprk7* (Fig. 4) and *Stk23*, and the protein kinase-A regulatory subunit *Prkra1* are plausibly involved in entrainment, outputs, or regulation of circadian period, as known for other kinases¹⁸. *Pbif* (Fig. 4) and *Tnfaip2* code for secreted factors that are plausible transmitters of circadian information

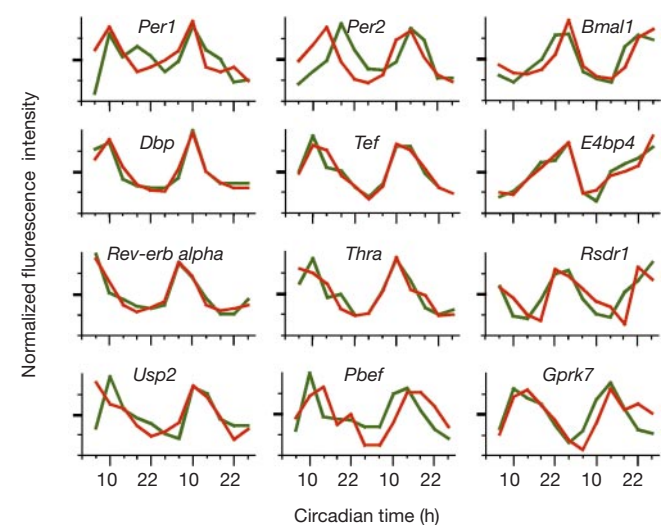


Figure 4 Individual circadian expression profiles of selected genes from the set common to liver and heart. Liver profiles are in green; heart profiles are in red. Gene symbols/names are indicated in Fig. 3.

within tissues or humorally to distant sites.

Five genes in the common set are regulated by factors with a diurnal or circadian rhythm in the circulation, of which glucocorticoids³ and retinoids⁴ have been proposed to act as entrainment signals for peripheral clocks. The gene *Thra* (Fig. 4), coding for the thyroid hormone receptor-α, is itself induced by the thyroid hormones T3 and T4 (ref. 19), and the phase of its rhythmic expression agrees with the humoral circadian rhythm of T3 and T4 in rat²⁰. Transcription of *Rev-erb-α* is inhibited by glucocorticoids¹⁶, and the timing of the decline in its expression agrees well

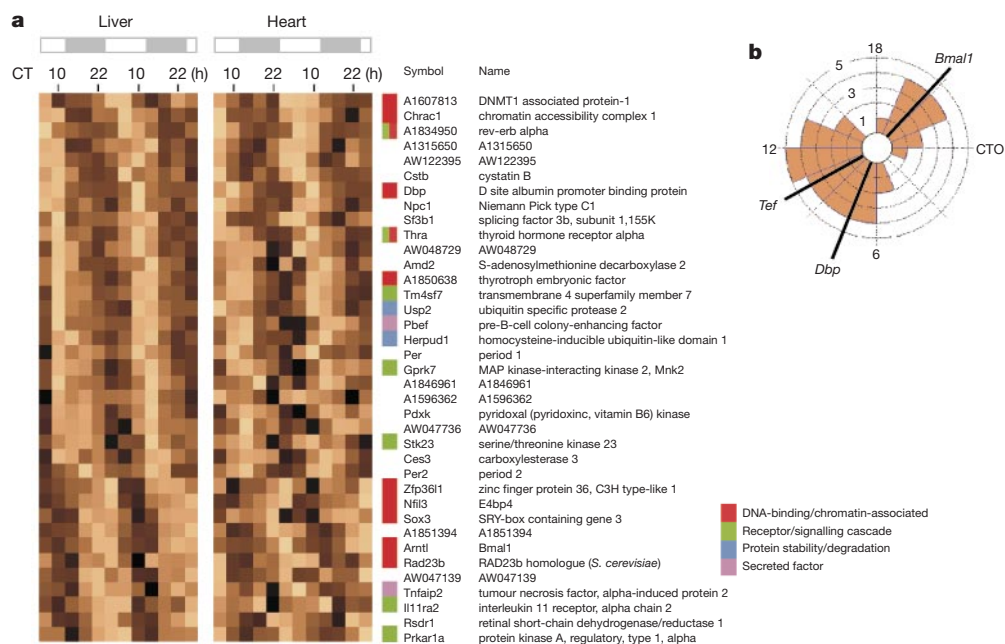


Figure 3 Temporal expression profiles of genes showing circadian regulation in both liver and heart. **a**, Profiles as in Fig. 1, ordered by phase as determined for liver. Annotation for selected genes is indicated by the colour code in the key; gene symbols and names are

from LocusLink. Uncharacterized genes are designated by GenBank accession numbers. **b**, Phase histogram of gene expression profiles in **a**; as in Fig. 1.

with the rising phase of the cortisol rhythm in mice²¹. The gene *Rsd* (Fig. 4) is induced by retinoids²², and *Tnfai2*²³ and *Zfp36*²⁴ are induced by tumour necrosis factor- α (TNF- α), both of which circulate with a diurnal rhythm^{25,26}. As transcriptional regulators, *Thra*, *Rev-erb-a* and *Zfp36* are plausible mediators of entrainment to humoral circadian signals, respectively thyroid hormones, glucocorticoids, and TNF- α .

Our work addresses global aspects of circadian gene regulation in mammals. We show that *in vivo* circadian gene regulation within individual tissues is extensive and diverse, broadly similar to circadian gene regulation in whole *Arabidopsis* plants²⁷ and *Drosophila* heads^{28,29}. Our comparative analysis of circadian gene expression in liver and heart reveals unsuspected features of circadian organization in peripheral tissues. Although the extent of circadian gene regulation in the two tissues is similar and the molecular clocks within the two tissues are in phase, liver and heart show little overlap in the timing and content of circadian gene regulation. Nevertheless, a global analysis of function suggests that circadian gene expression impinges on a large set of related and overlapping processes in the two tissues. These results suggest that a common core clock mechanism regulates similar, extremely diverse physiological processes in liver and heart, but that different downstream signalling circuits and effector genes carry out this regulation in the two tissues.

Note added in proof: A similar comparative analysis of circadian gene expression in the liver and suprachiasmatic nucleus has been reported³⁰. □

Methods

Animals, tissue collection and microarray hybridization

Mice (C57/BL6; Charles River Laboratories) were killed by decapitation under dim white light (less than 1 lx), and livers and hearts were removed, immediately frozen in liquid nitrogen, and stored at -70°C . Total RNA was extracted from individual tissues using Trizol (Life Technologies) and selected for poly(A)⁺ RNA by one round of oligo-deoxythymidine selection (Oligotex, Qiagen). For each time point, 5 μg of poly(A)⁺ RNA or, in a few cases, 10 μg of total RNA from tissue from a single mouse was used to make biotin-labelled complementary RNA (cRNA), which was hybridized on an Affymetrix U74Av2 oligonucleotide array. cRNA generated from poly(A)⁺ RNA and total RNA performed equivalently. We carried out cRNA synthesis, hybridization and array scanning according to the Affymetrix GeneChip Expression Analysis Technical Manual. For identifying and annotating genes represented on the array, GenBank accession numbers were used to search Unigene and LocusLink databases.

Data analysis

Raw scanner image files were analysed with dChip⁸. For each gene, the dChip outputs ('raw expression values') for the first six time points (day 1) and the second six time points (day 2) were each set to a mean of 1 and a standard deviation of 0. We treated them separately because day 1 and day 2 samples were processed and hybridized to arrays separately. For each gene, the two normalized (n) data sets were then concatenated to obtain a 12-time-point expression profile (nday 1–nday 2).

Four parameters (P) were used to select for circadian expression patterns: (P1), autocorrelation of the concatenated 12 time points (nday 1–nday 2 versus nday 1–nday 2); (P2), correlation of each day with the other (nday 1–nday 1 versus nday 2–nday 2); (P3), a threshold for the maximum raw expression value within each day; and (P4), a threshold within each day for the variation of the raw expression value (determined from the standard deviation) divided by the mean raw expression value.

To optimize the parameters for selection of circadian expression patterns, we designated nine guide genes with documented circadian expression in both liver and heart (see text). Parameter selection was performed by moving along the coordinates of the four-dimensional parameter space, simultaneously minimizing the total number of genes selected while maximizing the number of guide genes selected. For estimation of noise within the selected data sets, we randomly permuted the 12 time points (nday 1–nday 2) and then assigned each gene an amplitude and a mean, randomly selected from the genes on the array. This procedure keeps the content of the data set constant while permuting the features operated on by each of the selection parameters. For a full account of the filtering procedures, see Supplementary Information.

Gene ontology maps

The comparative analysis of expression data using the Gene Ontology structured vocabulary was performed using the GoSurfer analysis tool, which will be described elsewhere (O.L. & K-F.S. *et al.*, manuscript in preparation). Briefly, GoSurfer creates unique identifiers for every path in a gene ontology graph and implements statistical and other tools for analysing path properties and relationships. The node coordinates of the tree displayed in Fig. 2 were created as follows. First, we extracted gene ontology identifiers associated with the genes in each of the sets or their human orthologues from LocusLink

(www.ncbi.nlm.nih.gov/LocusLink) or GO Annotations at the European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk). Next we matched these gene ontology identifiers with the gene ontology for biological process (http://www.geneontology.org), and transformed this subset of the ontology into a tree structure, where each node represents a gene ontology identifier and where the branches indicate their hierarchical relationship as defined by the gene ontology. For the graphical presentation we used an existing Matlab program (Trimtreelayout) to generate the tree backbone.

Received 28 February; accepted 2 April 2002.

Published online 21 April 2002, DOI 10.1038/nature744.

- Balsalobre, A., Damiola, F. & Schibler, U. A serum shock induces circadian gene expression in mammalian tissue culture cells. *Cell* **93**, 929–937 (1998).
- Yamazaki, S. *et al.* Resetting central and peripheral circadian oscillators in transgenic rats. *Science* **288**, 682–685 (2000).
- Balsalobre, A. *et al.* Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science* **289**, 2344–2347 (2000).
- McNamara, P. *et al.* Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock. *Cell* **105**, 877–889 (2001).
- Reppert, S. M. & Weaver, D. R. Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* **63**, 647–676 (2001).
- Ripperger, J. A. & Schibler, U. Circadian regulation of gene expression in animals. *Curr. Opin. Cell Biol.* **13**, 357–362 (2001).
- Lockhart, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
- Li, C. & Wong, W. H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl Acad. Sci. USA* **98**, 31–36 (2001).
- Claverie, J.-M. Gene number: what if there are only 30,000 human genes? *Science* **291**, 1255–1257 (2001).
- Murphy, W. J., Stanyon, R. & O'Brien, S. J. Evolution of mammalian genome organization inferred from comparative gene mapping. *Genome Biol.* (Review) **2**, 0005.1–0005.8 (2001).
- Kornmann, B., Preitner, N., Rifat, D., Fleury-Olela, F. & Schibler, U. Analysis of circadian liver gene expression by ADDER, a highly sensitive method for the display of differentially expressed mRNAs. *Nucleic Acids Res.* **29**, E51 (2001).
- Velculescu, V. E. *et al.* Analysis of human transcriptomes. *Nature Genet.* **23**, 387–388 (1999).
- Klein, D. C., Moore, R. Y. & Reppert, S. M. (eds) *Suprachiasmatic Nucleus: The Mind's Clock* (Oxford Univ. Press, New York, 1991).
- Grundschober, C. *et al.* Circadian regulation of diverse gene products revealed by mRNA expression profiling of synchronized fibroblasts. *J. Biol. Chem.* **276**, 46751–46758 (2001).
- Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
- Torra, I. P. *et al.* Circadian and glucocorticoid regulation of Rev-erb- α expression in liver. *Endocrinology* **141**, 3799–3806 (2000).
- Ripperger, J. A., Shearman, L. P., Reppert, S. M. & Schibler, U. CLOCK, an essential pacemaker component, controls expression of the circadian transcription factor DBP. *Genes Dev.* **14**, 679–689 (2000).
- Young, M. W. & Kay, S. A. Time zones: a comparative genetics of circadian clocks. *Nature Rev. Genet.* **2**, 702–715 (2001).
- Tata, J. R., Baker, B. S., Machuga, L., Rabelo, E. M. & Yamauchi, K. Autoinduction of nuclear receptor genes and its significance. *J. Steroid Biochem. Mol. Biol.* **46**, 105–119 (1993).
- Kalsbeek, A., Fliers, E., Franke, A. N., Wort, J. & Buijs, R. M. Functional connections between the suprachiasmatic nucleus and the thyroid gland as revealed by lesioning and viral tracing techniques in the rat. *Endocrinology* **141**, 3832–3841 (2000).
- Harris, H. J., Kotelevtsev, Y., Mullins, J. J., Seckl, J. R. & Holmes, M. C. Intracellular regulation of glucocorticoids by 11 β -hydroxysteroid dehydrogenase (11 β -HSD)-1 plays a key role in regulation of the hypothalamic-pituitary-adrenal axis: analysis of 11 β -HSD-1-deficient mice. *Endocrinology* **142**, 114–120 (2001).
- Chen, J., Maltby, K. M. & Miano, J. M. A novel retinoid-response gene set in vascular smooth muscle cells. *Biochem. Biophys. Res. Commun.* **281**, 475–482 (2001).
- Wolf, F. W. *et al.* B94, a primary response gene inducible by tumour necrosis factor- α , is expressed in developing hematopoietic tissues and the sperm acrosome. *J. Biol. Chem.* **269**, 3633–3640 (1994).
- Cooper, P., Potter, S., Mueck, B., Yousefi, S. & Jarai, G. Identification of genes induced by inflammatory cytokines in airway epithelium. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **280**, L841–L852 (2001).
- Buchan, P. *et al.* Repeated topical administration of all-trans-retinoic acid and plasma levels of retinoic acids in humans. *J. Am. Acad. Dermatol.* **30**, 428–434 (1994).
- Petrovsky, N. & Harrison, L. C. The chronobiology of human cytokine production. *Int. Rev. Immunol.* **16**, 635–649 (1998).
- Harmer, S. L. *et al.* Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000).
- Claridge-Chang, A. *et al.* Circadian regulation of gene expression systems in the *Drosophila* head. *Neuron* **32**, 657–671 (2001).
- McDonald, M. J. & Rosbash, M. Microarray analysis and organization of circadian gene expression in *Drosophila*. *Cell* **107**, 567–578 (2001).
- Panda, S. J. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* [online 2 April 2001] 10.1016/S0092867402007225 (2002).

Supplementary Information accompanies the paper on Nature's website (http://www.nature.com).

Acknowledgements

This work was supported by the Edward R. and Anne G. Lefler Center, Harvard Medical School and the National Eye Institute (C.J.W.), a Deutsche Forschungsgemeinschaft Postdoctoral Fellowship (K-F.S.), the National Institute of Child Health and Human

Development (F.C.D. and N.V.), and the National Human Genome Research Institute (W.H.W.). We thank S. Meng and the Harvard Center for Genomics Research for technical assistance and M.-C. Kao for programming.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.J.W. (e-mail: cweitz@hms.harvard.edu).

Glutamate-receptor-interacting protein GRIP1 directly steers kinesin to dendrites

Mitsutoshi Setou*, Dae-Hyung Seog*†, Yosuke Tanaka*, Yoshimitsu Kanai*, Yosuke Takei*, Masahiko Kawagishi* & Nobutaka Hirokawa*

* Department of Cell Biology and Anatomy, Graduate School of Medicine, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

† Department of Microbiology, College of Medicine, Inje University, Jingu, Busan 633-165, Korea

In cells, molecular motors operate in polarized sorting of molecules, although the steering mechanisms of motors remain elusive¹. In neurons, the kinesin motor² conducts vesicular transport such as the transport of synaptic vesicle components to axons³ and of neurotransmitter receptors to dendrites⁴, indicating that vesicles may have to drive the motor for the direction to be correct. Here we show that an AMPA (α -amino-3-hydroxy-5-methylisoxazole-4-propionate) receptor subunit—GluR2-interacting protein (GRIP1)—can directly interact and steer kinesin heavy chains to dendrites as a motor for AMPA receptors. As would be expected if this complex is functional, both gene targeting and dominant negative experiments of heavy chains of mouse kinesin showed abnormal localization of GRIP1. Moreover, expression of the kinesin-binding domain of GRIP1 resulted in accumulation of the endogenous kinesin predominantly in the somatodendritic area. This pattern was different from that generated by the overexpression of the kinesin-binding scaffold protein JSAP1 (JNK/SAPK-associated protein-1, also known as Mapk8ip3), which occurred predominantly in the somatoaxon area. These results indicate that directly binding proteins can determine the traffic direction of a motor protein.

The neuron, a good model of a polarized cell, is divided into two molecularly and functionally distinct domains: axons and dendrites. The precise targeting and localization of proteins within these domains are critical to every aspect of neuronal function^{1,5}. In neurons, conventional kinesin, which consists of two heavy chains (KHC: KIF5) and two light chains (KLC)^{1,2}, is a multifunctional transporter of both axonal cargo such as synapsin and GAP43 (ref. 3) and dendritic cargo such as messenger RNA⁶ and the AMPA receptor⁷ (Fig. 1a). To elucidate the mechanisms possibly regulating how kinesin moves toward axons and/or dendrites, we screened the binding regulators of kinesin in a yeast two-hybrid system⁷. By this screening, we identified the glutamate-receptor-interacting protein GRIP1 (refs 8, 9) as the strongest partner of a cargo-binding domain^{10,11} of kinesin heavy chain, and we further showed that binding proteins could regulate the polarity of kinesin traffic direction.

All KIF5 isoforms (KIF5A, KIF5B and KIF5C) contained the

minimal GRIP1-binding site (Fig. 1b). The binding site overlapped with the cargo-binding domain of the fungus kinesin, which lacks the light chains¹¹. The tails of other major neuronal KIFs, such as KIF1A, KIF1B β and KIF17 (which bound to the mLin10 PDZ domain in this assay⁷), did not bind to GRIP1. The affinity of direct KIF5–GRIP1 interaction *in vitro* by surface plasmon resonance analysis⁷ ($K_d = 1.9 \times 10^{-8}$ M) was in good agreement with that previously reported for kinesin tail binding to brain microsomes *in vitro*¹⁰. Therefore, we suggest that heavy chains of kinesin directly bind to GRIP1.

We next investigated whether kinesin transports GRIP1. We used 'kinesin-null' primary cultured extra-embryonic cells from KIF5b-

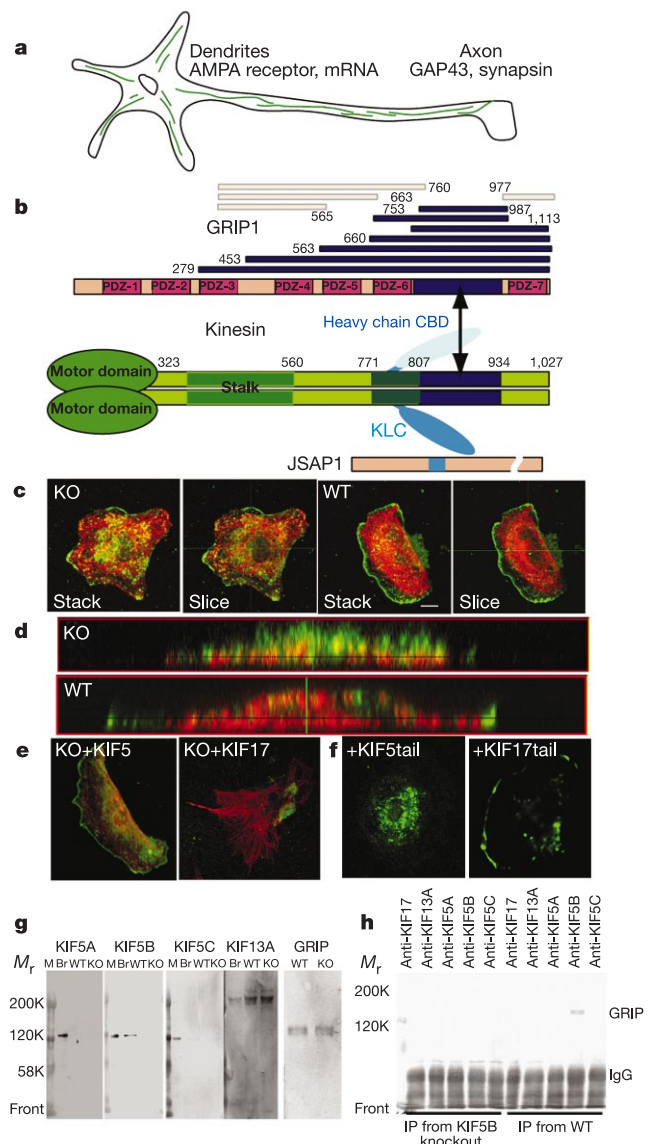


Figure 1 Kinesin binds to GRIP1, and is necessary for normal localization of GRIP1. **a**, Schematic diagram of a neuron and its reported kinesin cargo. **b**, Arrangement of GRIP1 and the kinesin binding domain. Blue boxes are the binding areas, with corresponding amino-acid positions indicated. Pale yellow boxes are areas of no binding. **c**, Aberrant localization of GRIP1 in cells lacking KIF5 (knockout cells, KO). WT, wild type. GRIP1 is green; microtubules are red. Slices are confocal plane images at one-third of cell height. Scale bar, 20 μ m. **d**, The x-z sections of the cells in **c**. **e**, KIF5, but not KIF17, rescued KIF5-knockout cells. **f**, KIF5 tail, but not KIF17 tail, disturbed the localization of GRIP1. **g**, KIFs and GRIP1 expression level in knockout and wild-type cells. **h**, GRIP1 binds KIF5B in these cells. IP, immunoprecipitate. M, molecular mass marker; Br, brain.

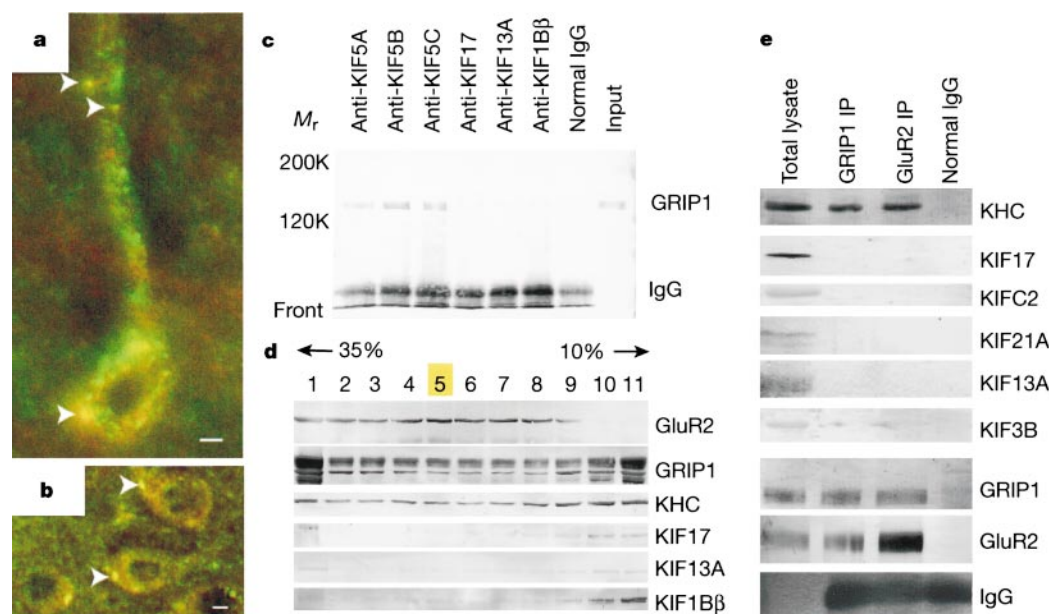


Figure 2 Kinesin-GRIP1-GluR2 complex *in vivo*. **a**, Kinesin (KIF5A, red) and GRIP1 (green). They co-localized in the soma and dendritic shafts (indicated with arrowheads), but not on the small scattered green dots (perhaps synaptic) or the red fibre that crosses (perhaps axonal). Scale bar, 2 μ m. **b**, The signal of KIF5A (in red) in the soma is related to Golgi (arrowheads), with β -COP (green). Scale bar, 2 μ m. **c**, Immunoprecipitation of

GRIP1 with KIFs. **d**, Kinesin-GRIP1-GluR2 transport vesicles. The middle fraction (5) was previously characterized as GRIP-GluR2-positive transporting vesicles¹⁸. KHC shows a wide, overlapping distribution. **e**, Immunoprecipitation of the kinesin-GRIP1-GluR2 complex.

knockout mice and control wild-type littermates¹². Both knockout and wild-type cells expressed GRIP1 equally (Fig. 1g), consistent with the previous description of embryonic GRIP expression¹³. If kinesin indeed transports GRIP1, then a lack of kinesin might result in a mislocalization of GRIP1 in these cells. Native GRIP1 clustered abnormally at the perinuclear region in knockout cells, which was different from the case in wild-type cells (Fig. 1c). In KIF5b-

knockout cells, most GRIP1 disappeared from the periphery, and appeared as relatively abundant intracellular clusters in the centre of a cell. In *x-z* cell sections, we could see a larger amount of intracellular stacked GRIP1 at the centre of knockout cells compared with wild-type (Fig. 1d) cells. This effect is specific because a lack of kinesin does not affect the bulk delivery systems of the endoplasmic reticulum, the Golgi, the *trans*-Golgi network (TGN)

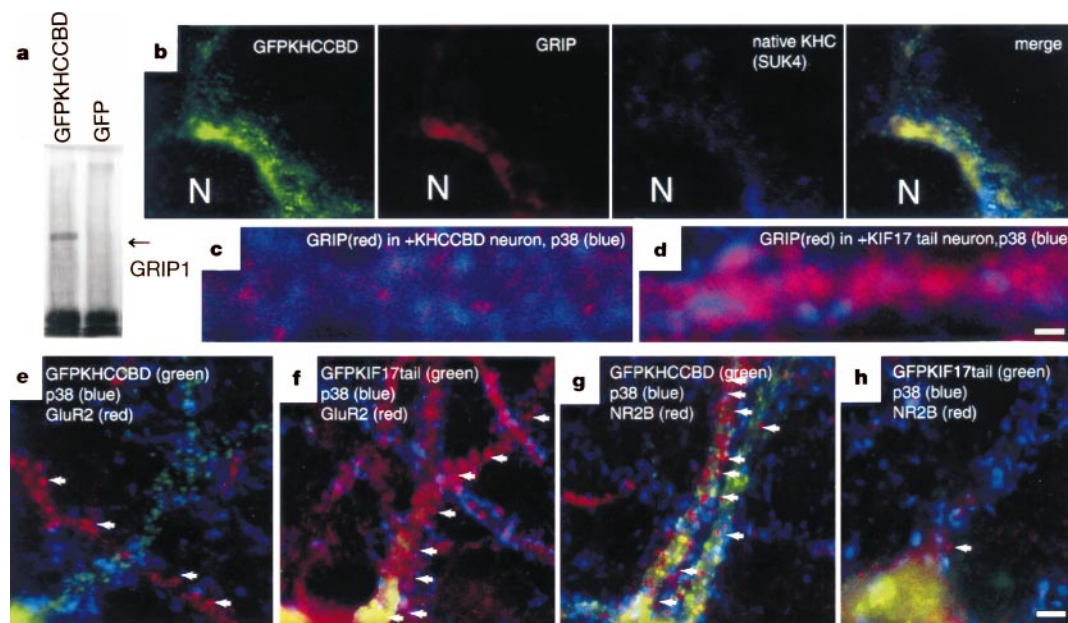


Figure 3 KHCCBD relocates GRIP1 and GluR2. **a**, GFP-KHCCBD binds native GRIP1 in neurons. **b**, GFP-KHCCBD and endogenous GRIP1 in the soma. The fourth panel is the merged image of the other three. Co-localization is observed around the nucleus (N). **c**, GRIP1 in distal dendrites. Synaptophysin (p38) labels presynapses. **d**, Control. Scale bar, 1 μ m. **e**, Dominant negative effects of GFP-KHCCBD. p38 is also slightly affected,

but GluR2 reduction is more significant. Arrows indicate the non-transfected dendrites, as an internal control. **f**, Arrows indicate normal GluR2 clusters in the dendrites of transfected neurons. **g**, GFP-KHCCBD does not affect NR2B. **h**, The GFP-KIF17 tail affects NR2B. The arrow indicates an NR2B cluster.

and the plasma membrane, nor does it affect localization of these molecules in the cells of the knockout mice¹². KIF5, but not KIF17, could rescue this phenotype (Fig. 1e). Exogenous KIF5 tail expression in wild-type cells also resulted in the abnormal clustering of GRIP1 (Fig. 1f). KIF5B antibody effectively immunoprecipitated GRIP1 from wild-type cells but not from knockout cells (Fig. 1h). Thus, kinesin can efficiently target GRIP1 to the cell periphery.

We then investigated the function of the kinesin–GRIP complex in neurons. Kinesin exists in both the somatodendritic area and axons¹⁴ and Golgi-related structures in the soma¹⁵ *in vivo*. GRIP1 localizes in Golgi-like structures in the soma, in vesicles in dendritic shafts, and in post-synaptic density (PSD)⁹. Kinesin and GRIP1 tend to co-localize in dendritic shafts and somatic Golgi-related structures in brain sections (Fig. 2a, b). GluR2 is known to exist in synapses and dendritic shafts¹⁶. A large population of the intracellular (including extrasynaptic) GluR2 forms a protein complex with GRIP1 (ref. 9). Moreover, scaffolding protein complexes can work as adapters for receptor traffic by motors^{7,17}. Therefore, we analysed the kinesin–GRIP–GluR2 complex *in vivo*. Antibodies against all the KIF5 isozymes immunoprecipitated GRIP from brain lysates (Fig. 2c). Kinesin exists in the subcellular fraction enriched in GluR2–GRIP-positive transporting vesicles¹⁸ (Fig. 2d). Antibodies to GRIP1 and GluR2 pulled kinesin down from this fraction (Fig. 2e). Thus, the GRIP1–GluR2 complex is one of the binding partners of kinesin in dendrites.

Having found that GRIP1 is a kinesin-binding scaffolding protein for GluR2 *in vivo*, we investigated whether the KIF5–GRIP complex transports GluR2, using a KIF5 dominant negative construct in neurons. We investigated the effect in cultured neurons of a construct (green fluorescent protein (GFP)–KHCCBD KIF5A807–1027) that contains the GRIP1-binding site, but lacks the KLC-binding site. When we immunoprecipitated the GRIP1–GFP–KHCCBD complex with an anti-GFP antibody, a significant

amount of GRIP1 was detected in the precipitate from cells expressing GFP–KHCCBD (Fig. 3a), as expected. GRIP1 and GFP–KHCCBD co-localized in the soma (Fig. 3b). The GFP–KHCCBD construct competed with the intrinsic KIF5–GRIP1 complex (Fig. 3b), and reduced the amounts of GRIP1 (Fig. 3c) and GluR2 in the synapses (Fig. 3e). Control KIF17 tail expression did not abolish GluR2 in the synapses (Fig. 3f). KIF5 dominant negative constructs did not significantly affect NMDA (*N*-methyl-D-aspartate) receptors in these cells (Fig. 3g), whereas KIF17 dominant negative constructs (Fig. 3h) did affect these receptors. Thus we showed that GluR2 is a passenger of the kinesin–GRIP1 transport machinery.

We wondered why kinesin does not transport the GluR2–GRIP complex into the axons. We had regarded kinesin as a major motor for axonal transport supplying materials into axons that are more than 99% of the volume of mature mammalian neurons *in vivo*. So we further dissected the molecular details of the GluR2–GRIP1–kinesin complex. The minimal kinesin-binding domains of GRIP1 (753–987) should be free from retention or reactivation with other possible binding partners (this construct lacks the binding sites for GluR2, self-multimerization sites and other potential regulator RasGEF-binding sites¹⁹, and did not bind endogenous GRIP1; Fig. 4m). This construct of GRIP1 delocalized kinesin predominantly in the somatodendritic area, but not significantly in axons (86% of the GRIP1 fragment signal co-localized with kinesin. Some 27% of the GRIP fragment and 39% of kinesin were located in dendrites (Fig. 4a–d). In contrast, the minimal binding domain of JSAP1 (mouse Syd2, JIP3), which functions as a binding partner of kinesin via KLC^{17,20,21}, delocalized kinesin predominantly in axons and soma (Fig. 4i–l) as a good control (92% of the JSAP1 fragment signal co-localized with kinesin; 4% of JSAP1 and 4% of kinesin were located at dendrites). Microtubules have their positive ends directed to the distal end in both axons and distal dendrites, but

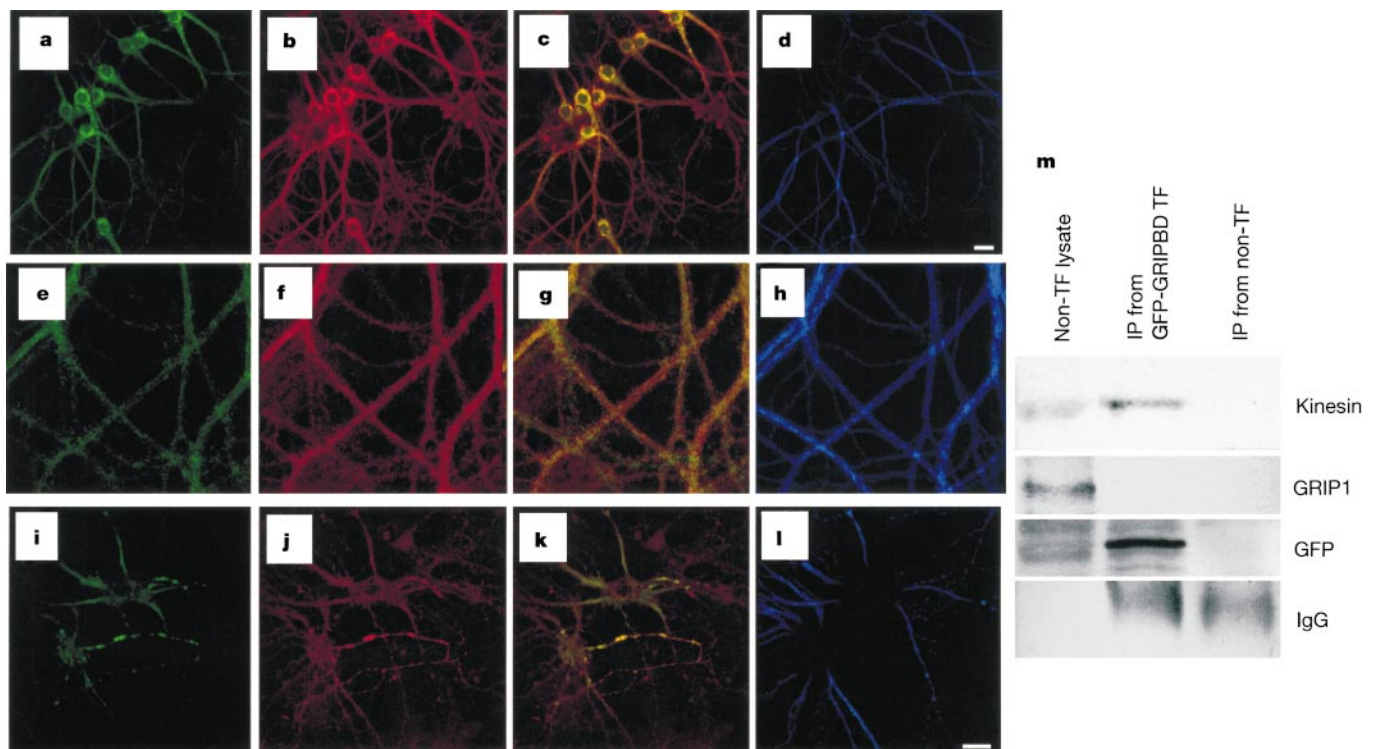


Figure 4 The GRIP1 kinesin-binding site recruits kinesin to the soma and dendrites. **a**, The Kinesin-binding domain of GRIP1–GFP (green). **b**, Intrinsic kinesin (red). **c**, GRIP1 and kinesin (merged, yellow) in the somatodendritic area. **d**, Dendrites stained by anti-MAP2 (blue). Scale bar, 10 μ m. **e–h**, Higher magnification of **a–d**, respectively. **i**, JSAP1

kinesin-binding domain (green). **j**, Kinesin (red). The JSAP1 kinesin-binding domain predominantly accumulates intrinsic kinesin in axon-like processes. **k**, Merged images of **i** and **j**. **l**, Dendritic marker map2 (blue). **m**, The GFP-minimal kinesin-binding domain of GRIP did not bind to native GRIP. TF, transfected.

microtubule polarity is mixed in proximal dendrites²². Thus, the 'plus-end-directed' motor kinesin can function in both directions^{5,7}. It has been confirmed that kinesin exists in axons, soma and dendrites under normal conditions^{14,23}, so the unbalanced ratio of the dendritic driver GRIP1 and axonal driver JSAP1 may cause the unbalanced redistribution of kinesin in this assay.

Findings on unique dendritic plus-end-directed motors such as KIF21B²⁴ have suggested that a 'smart motor' may differentiate dendrite-directed microtubules from axonal microtubules⁵. In this context, our results provide the evidence that a 'smart motor complex' may recognize the difference between the dendrite-directed and axon-directed microtubules; that is, smart units can be constructed of various complexes of the motor and its drivers. We examined the default microtubule preference of the kinesin motor engine without drivers in another study (T. Nakata and N.H., manuscript in preparation). Kinesin-binding proteins may also contribute to enhancing, converting or masking the properties of molecular engines. This concept allows the various cargoes to be carried by a limited number of motors to different directions at different times. Most of the other KHC binding partners previously reported, such as kinectin²⁵ or MyoVA²⁶, are mainly dendritic. Moreover, most of the KLC partners, such as JSAP^{21,17} and amyloid precursor protein²⁷, have been considered axon directed, implying that kinesin transport can be generally categorized as the KHC pathway for somatodendritic dominance, and the KLC pathway for somatoaxonal dominance. In summary, we have provided evidence that GRIP1 enables kinesin to transport dendritic proteins such as AMPA receptor subunit GluR2. □

Methods

Molecular biology constructs

KIF5A (808–1027) was used for yeast two-hybrid screening. Tail constructs of KIF5s, such as KIF5A (808–967), KIF5A (808–934), KIF5B (808–963 end), KIF5B (808–935), KIF5C (808–957 end), and KIF5C (808–935), were also positive for GRIP (279–1,113) binding. Other KIF constructs (KIF1A (400–end), KIF1B α (830–end), KIF1B β (830–end) and KIF17 (939–end)), or binding domain separation mutants (KIF5A (808–877), KIF5A (877–1,027), KIF5A (953–1,027), and KIF5B (808–883)) were negative for GRIP (279–1,113 end) binding, as negative controls. We identified JSAP1 as the strongest binding partner for KLC. We excluded the JSAP1 data from this paper because it has been reported recently²¹. Full-length GRIP1 and JSAP1 were cloned and sequenced from the mouse brain library⁷. GFP-fusion GRIP mutants were subcloned from pB42AD clones into the pEGFP2 vector (Clontech) using *EcoRI*–*XhoI* sites. Deleted or tagged constructs of KIFs were obtained from Hirokawa laboratory stocks. The Sindbis virus was constructed following the manufacturer's instruction (Sinrep system, Invitrogen). GFP, Myc and haemagglutinin tags were attached to GRIP1 or KIF amino-terminal ends, and virus titres were normalized using their titre on BHK cells.

Antibodies

Anti-KIF antibodies were obtained from Hirokawa laboratory stocks^{7,14}. SUK-4 was purchased from Covance, and H2 from Chemicon. The goat anti-GluR2 antibody was purchased from Santa-Cruz²⁸. The anti-GluR2 polyclonal antibody for immunoprecipitation was from Chemicon. The anti-NR2B polyclonal antibody was a gift from M. Mishina. The anti-Pan GRIP was a gift from M. Sheng, and the anti-GRIP1 and GRIP2 antibodies were gifts from R. L. Haganir. The anti-GRIP1 monoclonal antibody was purchased from Beckton Dickinson. The anti-MAP2 antibody, the anti- β -COP antibody, and anti-synaptophysin antibody were purchased from Sigma. The anti- α -tubulin monoclonal antibody was from Serotec. For secondary antibodies, Alexa 488, 568, goat anti-rabbit or mouse immunoglobulin- γ (Molecular Probe), Alexa 568 donkey anti-goat antibody (Molecular Probe), fluorescein isothiocyanate donkey anti-mouse antibody (Jackson Laboratories), anti-mouse Cy5 antibody (Amersham Pharmacia), and ALP-goat anti-mouse, goat anti-rabbit, and donkey anti-goat (Chapel) antibodies were used.

Immunoprecipitations

For brain analysis, about 10 μ g of affinity-purified rabbit anti-KIF antibodies precoupled with protein-A beads was mixed with 400 μ g of solubilized lysates at 4°C overnight (for KIF5s, antibodies against its rod were used). The beads were centrifuged for 2 min and washed with Triton-Tris-EDTA buffer three times. GRIP2 was weakly detected in the precipitates of kinesin, but the significance of this association was not investigated further. For subcellular fractionation and immunoprecipitation of the GRIP–GluR2 transport complex *in vivo*, we followed a previously described method¹⁸ without major modifications.

Primary culture analysis

Extra-embryonic cell culture and rescue were carried out as previously described¹². Biochemical analysis of dominant negative constructs was from mid-density-cultured rat

brain neurons²⁹ in 10-cm dishes for 4 weeks. Results of morphological analysis with respect to the effects of KIF5 dominant negative constructs on GRIP1 and GluR2 in neurons^{5,7} were presented. Images were processed with NIH image for numerical analysis¹². The co-localization of the molecules analysed was determined by dividing the pixel number with both signals by that of a single signal filtered into a normalized bitmap at a median threshold. The titres of the viruses were adjusted to the level where about 70% of the cells showed GFP expression.

Received 10 December 2001; accepted 25 March 2002.

Published online 14 April 2002, DOI 10.1038/nature743.

- Hirokawa, N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279**, 519–526 (1998).
- Vale, R. D., Reese, T. S. & Sheetz, M. P. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* **42**, 39–50 (1985).
- Ferreira, A., Niclas, J., Vale, R. D., Banker, G. & Kosik, K. S. Suppression of kinesin expression in cultured hippocampal neurons using antisense oligonucleotides. *J. Cell Biol.* **117**, 595–606 (1992).
- Kim, C. H. & Lisman, J. E. A labile component of AMPA receptor-mediated synaptic transmission is dependent on microtubule motors, actin, and N-ethylmaleimide-sensitive factor. *J. Neurosci.* **21**, 4188–4194 (2001).
- Burack, M. A., Silverman, M. A. & Banker, G. The role of selective transport in neuronal protein sorting. *Neuron* **26**, 465–472 (2000).
- Severt, W. L. *et al.* The suppression of testis-brain RNA binding protein and kinesin heavy chain disrupts mRNA sorting in dendrites. *J. Cell Sci.* **112**, 3691–3702 (1999).
- Setou, M., Nakagawa, T., Seog, D. H. & Hirokawa, N. Kinesin superfamily motor protein KIF17 and μ Lin-10 in NMDA receptor-containing vesicle transport. *Science* **288**, 1796–1802 (2000).
- Dong, H. *et al.* GRIP: a synaptic PDZ domain-containing protein that interacts with AMPA receptors. *Nature* **386**, 279–284 (1997).
- Wyszynski, M. *et al.* Association of AMPA receptors with a subset of glutamate receptor-interacting protein *in vivo*. *J. Neurosci.* **19**, 6528–6537 (1999).
- Skoufias, D. A., Cole, D. G., Wedaman, K. P. & Scholey, J. M. The carboxyl-terminal domain of kinesin heavy chain is important for membrane binding. *J. Biol. Chem.* **269**, 1477–1485 (1994).
- Seiler, S. *et al.* Cargo binding and regulatory sites in the tail of fungal conventional kinesin. *Nature Cell Biol.* **2**, 333–338 (2000).
- Tanaka, Y. *et al.* Targeted disruption of mouse conventional kinesin heavy chain, kif5B, results in abnormal perinuclear clustering of mitochondria. *Cell* **93**, 1147–1158 (1998).
- Bruckner, K. *et al.* EphrinB ligands recruit GRIP family PDZ adaptor proteins into raft membrane microdomains. *Neuron* **22**, 511–524 (1999).
- Kanai, Y. *et al.* KIF5C, a novel neuronal kinesin enriched in motor neurons. *J. Neurosci.* **20**, 6374–6384 (2000).
- Rahman, A., Kamal, A., Roberts, E. A. & Goldstein, L. S. Defective kinesin heavy chain behaviour in mouse kinesin light chain mutants. *J. Cell Biol.* **146**, 1277–1288 (1999).
- Rubio, M. E. & Wenthold, R. J. Differential distribution of intracellular glutamate receptors in dendrites. *J. Neurosci.* **19**, 5549–5562 (1999).
- Verhey, K. J. *et al.* Cargo of kinesin identified as JIP scaffolding proteins and associated signalling molecules. *J. Cell Biol.* **152**, 959–970 (2001).
- Lee, S. H., Valtschanoff, J. G., Kharazia, V. N., Weinberg, R. & Sheng, M. Biochemical and morphological characterization of an intracellular membrane compartment containing AMPA receptors. *Neuropharmacology* **41**, 680–692 (2001).
- Ye, B. *et al.* GRASP-1: a neuronal RasGEF associated with the AMPA receptor/GRIP complex. *Neuron* **26**, 603–617 (2000).
- Ito, M. *et al.* JSAP1, a novel jun N-terminal protein kinase (JNK)-binding protein that functions as a Scaffold factor in the JNK signalling pathway. *Mol. Cell Biol.* **19**, 7539–7548 (1999).
- Bowman, A. B. *et al.* Kinesin-dependent axonal transport is mediated by the sundry driver (SYD) protein. *Cell* **103**, 583–594 (2000).
- Baas, P. W., Deitch, J. S., Black, M. M. & Banker, G. A. Polarity orientation of microtubules in hippocampal neurons: uniformity in the axon and nonuniformity in the dendrite. *Proc. Natl Acad. Sci. USA* **85**, 8335–8339 (1988).
- Niclas, J., Navone, F., Hom-Booher, N. & Vale, R. D. Cloning and localization of a conventional kinesin motor expressed exclusively in neurons. *Neuron* **12**, 1059–1072 (1994).
- Marszalek, J. R., Weiner, J. A., Farlow, S. J., Chun, J. & Goldstein, L. S. Novel dendritic kinesin sorting identified by different process targeting of two related kinesins: KIF21A and KIF21B. *J. Cell Biol.* **145**, 469–479 (1999).
- Toyoshima, I., Yu, H., Steuer, E. R. & Sheetz, M. P. Kinectin, a major kinesin-binding protein on ER. *J. Cell Biol.* **118**, 1121–1131 (1992).
- Huang, J. D. *et al.* Direct interaction of microtubule- and actin-based transport motors. *Nature* **397**, 267–270 (1999).
- Kamal, A., Stokin, G. B., Yang, Z., Xia, C. H. & Goldstein, L. S. Axonal transport of amyloid precursor protein is mediated by direct binding to the kinesin light chain subunit of kinesin-I. *Neuron* **28**, 449–459 (2001).
- Burette, A. *et al.* Differential cellular and subcellular localization of AMPA receptor-binding protein and glutamate receptor-interacting protein. *J. Neurosci.* **21**, 495–503 (2001).
- Brewer, G. J. Serum-free B27/neurobasal medium supports differentiated growth of neurons from the striatum, substantia nigra, septum, cerebral cortex, cerebellum, and dentate gyrus. *J. Neurosci. Res.* **42**, 674–683 (1995).

Acknowledgements

We thank R. L. Haganir for the anti-GRIP1 and anti-GRIP2 antibodies, M. Mishina for anti-NR2B antibody, and M. Sheng for an anti-PanGRIP antibody. We thank Y. Hayashi for advice and for the Sindbis virus construction. We also thank Y. Hata, T. Nakagawa and M. Sheng for discussions; and H. Sato, H. Fukuda, T. Matsuki and M. Sugaya for technical assistance. Finally, we thank other members of the Hirokawa laboratory, especially T. Nakata, for technical assistance, discussions, and advice throughout the conducting of experiments. This work is supported by a Center of Excellence Grant-in-Aid from the Ministry of Education, Science, Sports, Culture and Technology of Japan to N. Hirokawa.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.H.
(e-mail: Hirokawa@m.u-tokyo.ac.jp).

Rapid regulation of steroidogenesis by mitochondrial protein import

Himangshu S. Bose*, Vishwanath R. Lingappa†‡ & Walter L. Miller*§

Departments of * Pediatrics, † Physiology and ‡ Medicine, and § The Metabolic Research Unit, University of California, San Francisco, California 94143-0978, USA

Most mitochondrial proteins are synthesized on cytoplasmic ribosomes and imported into mitochondria^{1–3}. The imported proteins are directed to one of four submitochondrial compartments—the outer mitochondrial membrane, the inner mitochondrial membrane, the intramembraneous space, or the matrix—where the protein then functions. Here we show that the steroidogenic acute regulatory protein (StAR), a mitochondrial protein required for stress responses, reproduction, and sexual differentiation of male fetuses^{4–7}, exerts its activity transiently at the outer mitochondrial membrane rather than at its final resting place in the matrix. We also show that its residence time at this outer membrane and its activity are regulated by its speed of mitochondrial import. This may be the first example of a

mitochondrial protein exerting its biological activity in a compartment other than that to which it is finally targeted. This system enables steroidogenic cells to initiate and terminate massive levels of steroidogenesis within a few minutes, permitting the rapid regulation of serum steroid hormone concentrations.

Steroid hormones, which are essential for life in all vertebrates, are made from cholesterol. The first and rate-limiting step in steroidogenesis is conversion of cholesterol to pregnenolone by the mitochondrial cholesterol side-chain cleavage enzyme, P450_{scc}, located on the matrix side of the inner mitochondrial membrane (IMM)⁸. Steroidogenic cells store minimal amounts of hormone; the regulation of steroid secretion and serum concentrations is primarily at the level of steroid hormone synthesis. Because of the powerful actions of steroid hormones and the toxicities resulting from their overproduction, vertebrates have evolved mechanisms for the rapid induction and termination of their synthesis. Tropic hormones stimulate transcription of genes encoding steroidogenic enzymes, providing a slow regulation that determines the identity and capacity of steroidogenic cells. However, the ability to induce steroidogenesis 10–100-fold within minutes and terminate this synthesis rapidly is regulated at the level of cholesterol flow from the outer mitochondrial membrane (OMM) to the IMM⁶. This acute response requires StAR^{4,5}, which can confer an acute steroidogenic response to non-steroidogenic cells co-transfected with the P450_{scc} system⁵. StAR mutations cause potentially lethal congenital lipoid adrenal hyperplasia, in which virtually no steroids are produced⁵. This disease develops in two stages, corresponding to the loss of the acute and then the chronic steroidogenic response to tropic stimuli⁷.

The mechanism and site of action of StAR are controversial. StAR is produced as a pre-protein of relative molecular mass 37,000

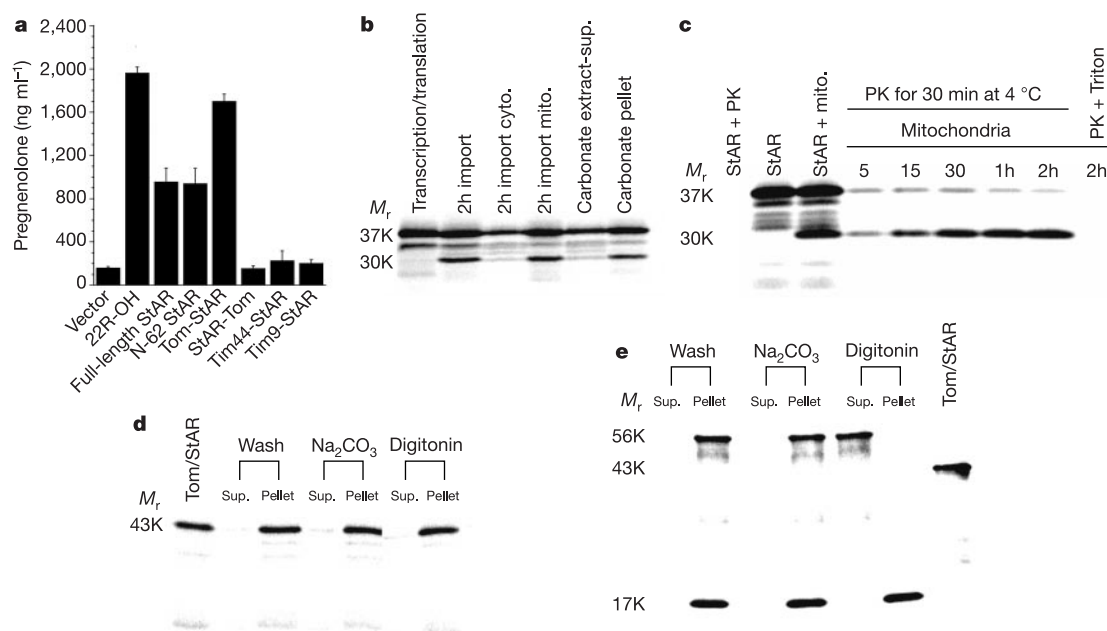


Figure 1 Affixing StAR to the outer mitochondrial membrane increases activity. **a**, Steroidogenic activity (accumulated pregnenolone secreted after 48 h), of COS-1 cells co-transfected with the F2 fusion of the cholesterol side-chain cleavage system and with vectors expressing the indicated constructions. 22R-OH, steroidogenesis from control cells incubated with 22-R-OH cholesterol; all other data show steroidogenesis from endogenous cellular cholesterol. **b**, Import of [³⁵S]StAR into MA-10 mitochondria. After 2 h some of the M_r 37K cell-free translation product is incorporated into the membrane and cleaved to 30K, which remained membrane-associated. **c**, Import of StAR into MA-10

mitochondria. Extramitochondrial 37K StAR is partially sensitive to proteinase K (PK), but intramitochondrial 30K StAR accumulates with time and remains protease-protected without Triton X-100. **d**, Import of Tom/StAR. Tom/StAR of M_r 43K associates with mitochondria but is not imported; carbonate and digitonin extractions show association with OMM (cyto., cytoplasm; mito., mitochondria; sup., supernatant). **e**, Digitonin selectively solubilizes the IMM. MA-10 mitochondria were prepared, washed and serially extracted with carbonate and digitonin; the indicated fractions were separated by gel electrophoresis and immunoblotted with a mixture of antisera to human P450_{scc} and to human Tom20.

(M_r 37K) with a mitochondrial leader sequence that is cleaved to yield stable intramitochondrial StAR of M_r 30K. However, N-62 StAR lacking 62 amino-terminal residues remains in the cytoplasm and has full steroidogenic activity⁹, but can also transfer cholesterol to other membranes such as the endoplasmic reticulum¹⁰. Thus,

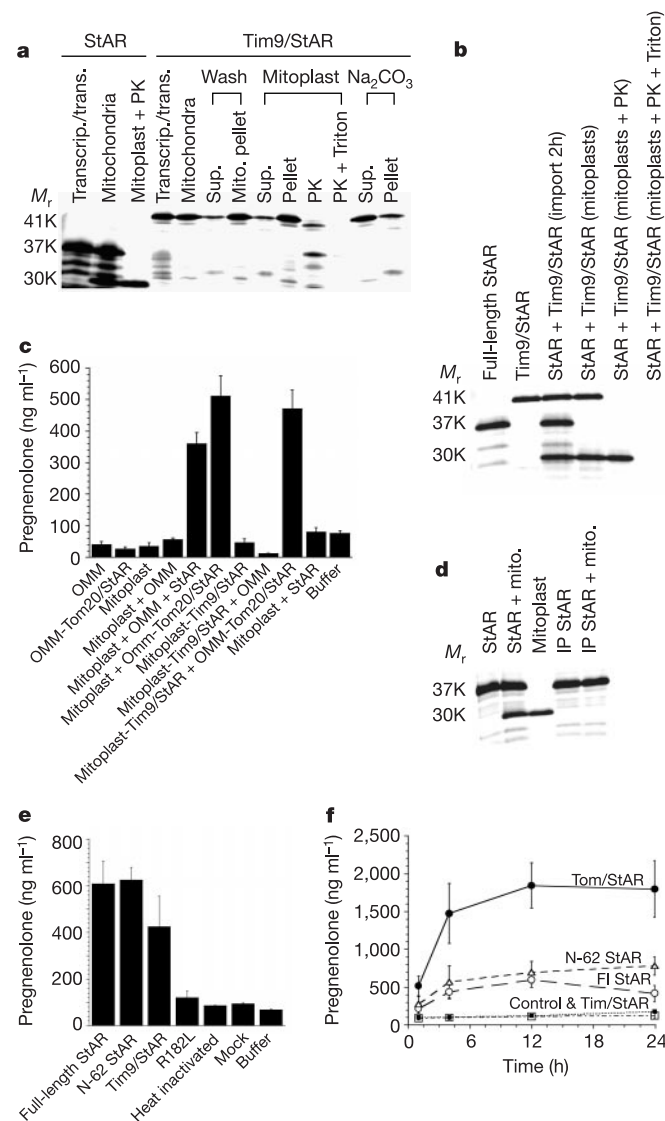


Figure 2 StAR is inactive in the intramembraneous space. **a**, Import of Tim9/StAR. Tim9/StAR of M_r 41K is associated with mitochondria and with mitoplasts, but extraction of mitoplasts with Na₂CO₃ or their incubation with protease shows that Tim9/StAR resides on the IMS side of the IMM. **b**, The presence of Tim9/StAR in the IMS does not disrupt the OMM or the import and cleavage of full-length StAR. **c**, Reconstitution of mitochondrial activity from OMM and mitoplast fractions. Fractions containing StAR constructs are indicated by dashes, mixing of separate components is indicated by plus signs; the system is active only with Tom/StAR affixed to the OMM or with exogenously added StAR. **d**, StAR expressed in the transcription/translation system is imported into mitochondria and cleaved to a 30K form in mitoplasts, but when StAR is immunoprecipitated (IP) from the transcription/translation system it cannot be incorporated into mitochondria. **e**, Tim9/StAR immunoprecipitated from the transcription/translation system is nonetheless active when added to MA-10 mitochondria *in vitro*. Controls include immunoprecipitation of the transcription/translation system expressing full-length StAR, N-62 StAR or expressing no StAR, the StAR mutant R182L (inactive *in vivo*⁷ and *in vitro*²⁰), buffer alone, use of heat treated mitochondria and use of cell-free transcription/translation system without added plasmid (mock). **f**, Steroidogenesis of MA-10 mitochondria incubated for various times with non-radioactive cell-free transcription/translation system without added vector (control) and with equal amounts of StAR, N-62 StAR, Tom/StAR and Tim9/StAR, as assayed by incorporation of [³⁵S]methionine in a parallel experiment.

StAR's mitochondrial leader peptide confines StAR's activity to mitochondria, and prevents it from transferring cholesterol to other membranes¹¹. These observations suggest that entry into the mitochondrion is not necessary for StAR's activity, and that StAR acts on or in the OMM¹¹. The 63–193 domain of StAR is protease-resistant, suggesting tight folding, slowing StAR's mitochondrial import, prolonging contact with the OMM, and increasing activity¹². N-62 StAR undergoes conformational changes to a molten globule at pH 3.5–4.0, both in solution and in association with synthetic unilamellar lipid membranes^{12–14}; this conformational change appears to foster association with the OMM, causing cholesterol binding and transfer to the IMM. However, the crystal structure of the related protein N-216 MLN64 features a hydrophobic pocket that should bind one molecule of cholesterol, suggesting that StAR functions as a cholesterol shuttling molecule within the intramembraneous space (IMS)¹⁵. This is inconsistent with the view that N-62

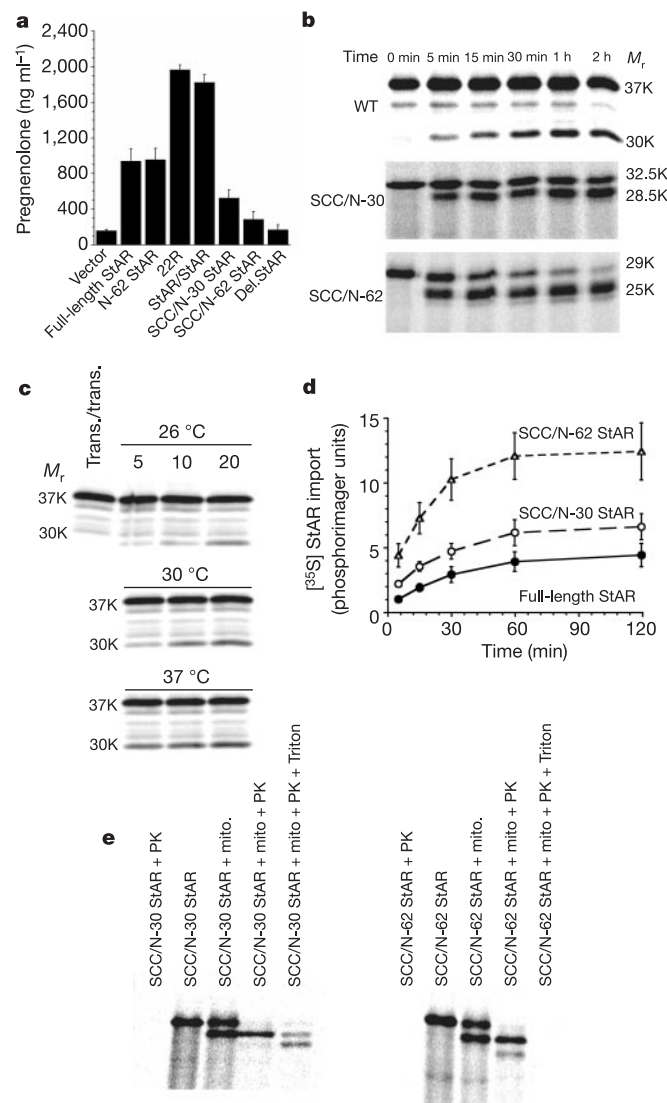


Figure 3 Association of StAR activity with mitochondrial import. **a**, Steroidogenesis of COS-1 cells co-transfected with F2 and the indicated vectors. Fusion of 1–193 StAR to 63–285 StAR (StAR/StAR), yields maximal activity (equal to 22R-OH-cholesterol); Scc/N-30 StAR, Scc/N-62 StAR, and del. StAR reduce activity. **b**, Time-course of import and cleavage (in minutes or hours). WT, wild type. **c**, Effect of incubation temperature on import kinetics (in minutes) of full-length StAR into MA-10 mitochondria. **d**, Kinetic analysis of data in **b**; data are mean $\pm$ s.d. from three independent experiments. **e**, Scc/N-30 StAR and Scc/N-62 StAR are imported and cleaved to intramitochondrial 30K protein that is protected from protease digestion in the absence of detergent.

StAR functions on the OMM or assumes a molten globule configuration¹⁵. To discriminate between these two conflicting models, we sought to determine whether StAR acts on the OMM, and, if so, whether its rate of mitochondrial entry and OMM residency time correlated with its activity.

During import into mitochondria, mitochondrial proteins are processed by a complex machinery consisting of Tom (translocase, outer membrane) and Tim (translocase, inner membrane) proteins^{1–3}. To determine the site of StAR's action, we immobilized StAR in various mitochondrial compartments by transfecting COS-1 cells with plasmids encoding StAR fused to Tom or Tim proteins; see Methods for details. The cells were co-transfected with a vector (F2) for a catalytically self-sufficient fusion of the cholesterol side chain cleavage enzyme system (H₂N-P450_{scc}-adrenodoxin reductase-adrenodoxin-COOH)¹⁶, which converts cholesterol to pregnenolone, providing an index of StAR activity. Tom20 has 50 N-terminal residues embedded in the OMM and 95 residues in the cytoplasm¹⁷. Expression of Tom20 with N-62 StAR fused to its carboxy terminus (Tom/StAR) affixed StAR to the cytoplasmic side of the OMM. Expression of Tim9, which is normally confined to the IMS³, fused to N-30 StAR (Tim9/StAR) localized StAR to the IMS. Expression of Tim44, which lies on the matrix side of the IMM, fused to N-62 StAR (Tim44/StAR), targeted StAR to the matrix.

Western blots with antiserum to human StAR indicated that each StAR construct was expressed at equivalent levels (not shown). Expression of F2 alone showed low levels of StAR-independent steroidogenesis, as described^{6,7}. Incubation with 22R-OH-cholesterol, which bypasses the action of StAR^{18,19}, showed maximal steroidogenic capacity ~20-fold greater than the StAR-independent level (Fig. 1a). Co-expression of F2 and full-length (1–285)

StAR or N-62 StAR showed a 6-fold increase in steroidogenesis over the StAR-independent level seen with F2 alone, or about half the maximal capacity of the cells seen with 22R-OH-cholesterol²⁰. Co-expression of Tom/StAR and F2 achieved maximal steroidogenesis equal to that seen with exogenously added 22R-OH-cholesterol; in contrast, StAR/Tom, Tim44/StAR and Tim9/StAR did not increase steroidogenesis (Fig. 1a). Thus localizing StAR to the OMM substantially increased activity, but StAR localized to the IMS or matrix was inactive.

Mitochondrial import assays²¹ show that Tom/StAR is associated with the OMM. Full-length [³⁵S]StAR prepared *in vitro* by a linked transcription/translation system was imported into mouse Leydig MA-10 cell mitochondria, and cleaved to a form of *M_r* 30K associated with the mitochondrial pellet. Extraction with Na₂CO₃, which disrupts protein–protein interactions but not lipid–protein interactions^{22,23}, showed that the imported 30K StAR form and some of the 37K protein were associated with lipid membranes (Fig. 1b). Extramitochondrial 37K StAR is only partially sensitive to proteinase K, suggesting that it is tightly associated with the OMM; intramitochondrial 30K StAR is wholly insensitive to proteinase K, and membrane solubilization with Triton X-100 renders both forms protease-sensitive (Fig. 1c). [³⁵S]Tom/StAR was also associated with mitochondria, as extraction with Na₂CO₃ did not dissociate [³⁵S]Tom/StAR from OMM (Fig. 1d). Extraction of the post-carbonate pellet with digitonin dissolves the IMM but not the OMM^{23,24} (Fig. 1e), but digitonin did not solubilize [³⁵S]Tom/StAR (Fig. 1d). Thus Tom/StAR is incorporated into the outer, but not the inner, mitochondrial membrane, where it exhibits substantially increased StAR activity, and this incorporation requires interaction with membrane lipids.

Mitochondrial import followed by stripping away the OMM to produce mitoplasts showed that 37K StAR was not retained but 30K StAR was, validating the mitoplast preparation. Uncleaved Tim9/StAR (*M_r* 41K) was associated with mitoplasts, but was sensitive to proteinase K or could be released to the supernatant by Na₂CO₃, showing it lies in the IMS (Fig. 2a). The presence of Tim9/StAR in the IMS did not disrupt the integrity of the OMM or mitochondrial function, as Tim9/StAR did not disrupt the import (Fig. 2b) or activity (not shown) of full-length StAR. When OMM and mitoplasts are separated, each component is steroidogenically inactive, irrespective of whether it contains a StAR construct. When mitochondria are reconstituted they are active with exogenously added StAR or when the OMM fraction contains Tom20/StAR; the presence or absence of Tim9/StAR in the mitoplast preparation contributes no activity, and exogenously added StAR is inactive with mitoplasts in the absence of OMM (Fig. 2c). Although Tim9/StAR was inactive in the IMS, it was active on the OMM. Tim9/StAR separated from the reticulocyte lysate system by immunoprecipitation could not be imported (Fig. 2d), but elicited activity on isolated mitochondria (Fig. 2e). Thus Tim9/StAR is inactive because of its IMS localization and is not inherently inactive from misfolding.

Measurement of steroidogenic activity of isolated mitochondria following StAR importation shows that the cell-free system faithfully recapitulates the key features of StAR biogenesis and activity observed *in vivo* (Fig. 2f). Mitochondria incubated with control cell-free transcription/translation system or that expressing Tim9/StAR yielded minimal steroidogenesis, but expression of either full-length StAR or N-62 StAR increased steroidogenesis ~6-fold, and expression of Tom/StAR increased activity 15–20 fold. Thus Tom/StAR is substantially more potent than full-length or N-62 StAR, both in cells (Fig. 1a) and in isolated mitochondria. Although some investigators categorize active and inactive forms of StAR on the basis of size (30K versus 37K protein), our data show that StAR's activity is determined by its mitochondrial location rather than by its size. This is consistent with recent data indicating that only newly synthesized StAR is active²⁵. Thus Fig. 1 shows that StAR is active on the OMM, and Fig. 2 shows it can not act in the IMS.

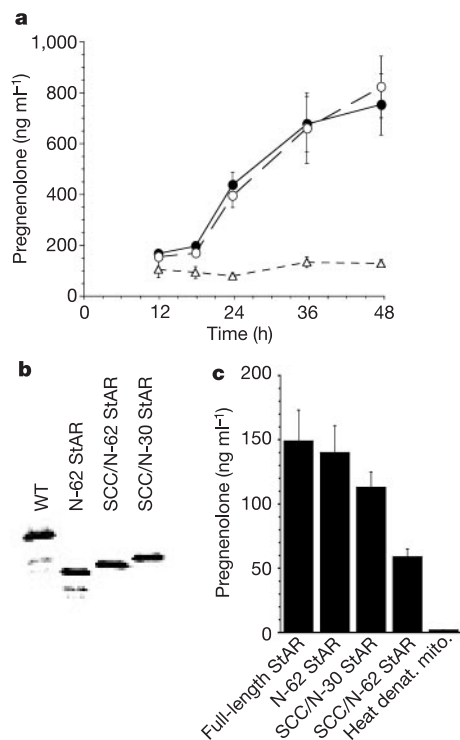


Figure 4 Full-length and N-62 StAR are equally active. **a**, Equimolar amounts of empty vector (triangles), vector expressing full-length StAR (filled circles) and N-62 StAR (open circles) were transfected into COS-1 cells and the accumulated pregnenolone in the culture medium was measured at the times shown. **b**, Immunoblot of equimolar amounts of full-length, N-62, SCC/N-62 and SCC/N-30 StAR. **c**, Activity of MA-10 cell mitochondria incubated for 1 h with equal masses of each protein prepared by cell-free transcription/translation.

As StAR activity is confined to the OMM, slowing import should increase activity. StAR residues 63–188 are protease-resistant and may slow StAR's import¹², so we added a second copy of 63–188 in front of 63–285 StAR (StAR/StAR). Conversely, to speed StAR's entry, we replaced StAR residues 1–30 (Scc/N-30 StAR) and residues 1–62 (Scc/N-62 StAR) with the 39-amino-acid leader of P450_{scc}, and we deleted residues 31–62 from wild-type StAR (del. StAR). When transfected into COS-1 cells co-transfected with F2, StAR/StAR achieved activity equal to that with 22R-OH-cholesterol, Scc/N-30 StAR had about half the activity of full-length or N-62 StAR, and Scc/N-62 StAR and del. StAR were equivalent to control (Fig. 3a). Thus constructs designed to slow mitochondrial entry increased activity and those designed to speed entry decreased activity. To estimate mitochondrial entry kinetics, we prepared each of these constructs by translation/transcription and incubated them with MA-10 cell mitochondria (Fig. 3b) at 26 °C, which slows mitochondrial import, facilitating comparison between constructs (Fig. 3c). Phosphorimager quantification of the uncleaved and the cleaved imported protein showed import speeds for Scc/N-62 StAR > Scc/N-30 StAR > full-length StAR (Fig. 3d). Both the Scc/N-30 StAR and Scc/N-62 StAR constructs were imported into mitochondria, as evidenced by their cleavage from a full-length form to a protease-resistant intramitochondrial 30K form (Fig. 3e). Thus StAR leader mutants that decreased activity speed mitochondrial import.

Others have suggested that the activity of N-62 StAR is an artefact due to overexpression in transfected cells¹⁵. When COS-1 cells are transfected with equimolar amounts of expression vectors for full-length or N-62 StAR, the steroidogenic activity conferred to the cells is equivalent at all times measured after transfection (Fig. 4a). To ensure that this reflects the activity of equivalent amounts of protein, the amounts of wild-type, N-62, Scc/N-30 and Scc/N-62 StAR produced by *in vitro* transcription/translation were estimated by immunoblotting compared to known amounts of bacterially expressed N-62 StAR²⁰ (Fig. 4b). When equal amounts of each protein were added to MA-10 cell mitochondria, full-length and N-62 StAR were equally active, Scc/N-30 StAR was slightly less so, and Scc/N-62 StAR had about half the activity of wild-type or N-62 StAR (Fig. 4c). Thus the activity of N-62 StAR expressed from transfected plasmids is equivalent to full-length StAR, and is not a pharmacological artefact.

These data show the following: (1) that StAR functions on or in the OMM, a location it maintains transiently; (2) that StAR's activity is proportional to its residency time on or in the OMM; (3) that the mitochondrial importation of StAR terminates its activity; and (4) that sequences in the first 62 amino acids of StAR specifically slow StAR's mitochondrial import compared to the leader of P450_{scc}. This indicates that StAR's activity is regulated by its rate of mitochondrial import. Because mitochondrial import is rapid, this unusual system permits rapid regulation of steroidogenesis in response to physiological needs, and establishes the rate of mitochondrial protein translocation as another site of physiological regulation. □

Methods

Construction of vectors

Tom20, Tim9 and Tim44 complementary DNAs were prepared from human adrenal RNA by PCR after reverse transcription of RNA (RT-PCR). N-62 StAR cDNA including an enterokinase linker fused to 63–285 StAR was derived from the bacterial expression vector¹⁶. Tom/StAR is Tom20 fused to 63–285 through GSDDDDK, which contains BamHI and enterokinase sites (Bam/EK linker). StAR/Tom is 1–285 StAR fused to Tom20 through EcoRI/EK linker (GFDDDDK). Tim9/StAR contains the 25-amino-acid mitochondrial leader of cytochrome c oxidase subunit IV²⁷ fused to Tim9 linked to N-30 StAR through the Bam/EK linker. Tim44/StAR is Tim44 fused to 63–285 StAR through the Bam/EK linker. StAR/StAR is StAR fused to 63–285 StAR with the Bam/EK linker. For Scc/N-30 and Scc/N-62 StAR, cDNA encoding amino acids 1–39 of human P450_{scc}²⁸ was fused directly to N-30 StAR or N-62 StAR. Del StAR is StAR residues 1–30 fused directly to 63–285. Constructs were verified by sequencing and expressed in the BamHI/EcoRI site of pCMV-Flag (Stratagene).

Cell culture

COS-1 cells were grown, transfected and assayed for pregnenolone production as described^{46,20}. Three independent transfections, each performed in triplicate, were done for each construct.

Protein import assays

StAR constructs were cloned into the BglII/EcoRI site of pSP6, transcribed with SP6 RNA polymerase (Promega) and translated with [³⁵S]methionine as described²⁹. Mitochondria, isolated from 5 × 10⁶ MA-10 or COS-1 cells by hypotonic shock and homogenization²⁰ were suspended in 100 µl of 125 mM sucrose, 1 mM ATP, 1 mM NADH, 50 mM KCl, 0.05 mM ADP, 2 mM DTT, 5 mM Na-succinate, 2 mM Mg(OAc)₂, 2 mM KH₂PO₄ and 10 mM HEPES buffer at pH 7.4. Mitochondria (50 µl) and 1 µl of the transcription/translation mix were incubated at 26 °C for various times; import was terminated with 1 µl 1 mM CCCP (Calbiochem). Some mitochondria were extracted with fresh 100 mM Na₂CO₃, pH 11.5, for 15 min at 4 °C and spun at 15,000g for 15 min at 4 °C to pellet membranes. Some membranes were incubated with 0.1% digitonin (Sigma) for 15 min at 4 °C and spun at 144,000g for 1 h at 4 °C to pellet OMM fragments. Mitoplast preparation: mitochondria were incubated with 1 U trypsin for 15 min at 4 °C, washed with 50 mM NaCl, 10 mM HEPES pH 6.5 containing 10 U soybean trypsin inhibitor and osmotically shocked in 10 mM HEPES pH 7.5 for 30 min at 4 °C (ref. 30). Analysis was by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and phosphorimager quantification (Molecular Dynamics, Storm 860).

Activities in isolated mitochondria

StAR constructs were bioassayed²⁰ with MA-10 mitochondria. Linked transcription/translation system was added directly to the mitochondria except when Tim9/StAR was immunoprecipitated with anti-StAR antiserum¹², collected with protein A-sepharose CL-4B (Pharmacia), washed, dissociated from the sepharose–antibody complex at pH 3.0, neutralized and added directly to mitochondria. Immunoassays of pregnenolone were performed after incubation at 37 °C for 1 h; immunoblotting was performed as described²⁰.

Received 24 January; accepted 28 February 2002.

- Schatz, G. & Dobberstein, B. Common principles of protein translocation across membranes. *Science* **271**, 1519–1526 (1996).
- Neupert, W. Protein import into mitochondria. *Annu. Rev. Biochem.* **66**, 863–917 (1997).
- Bauer, M. F. & Neupert, W. Import of proteins into mitochondria: A novel pathomechanism for progressive neurodegeneration. *J. Inher. Metab. Dis.* **24**, 166–180 (2001).
- Clark, B. J., Wells, J., King, S. R. & Stocco, D. M. The purification, cloning and expression of a novel luteinizing hormone-induced mitochondrial protein in MA-10 mouse Leydig tumour cells. Characterization of the steroidogenic acute regulatory protein (StAR). *J. Biol. Chem.* **269**, 28314–28322 (1994).
- Lin, D. *et al.* Role of steroidogenic acute regulatory protein in adrenal and gonadal steroidogenesis. *Science* **267**, 1828–1831 (1995).
- Stocco, D. M. & Clark, B. J. Regulation of the acute production of steroids in steroidogenic cells. *Endocr. Rev.* **17**, 221–244 (1996).
- Bose, H. S., Sugawara, T., Strauss, J. F. III & Miller, W. L. The pathophysiology and genetics of congenital lipid adrenal hyperplasia. *N. Engl. J. Med.* **335**, 1870–1878 (1996).
- Miller, W. L. Molecular biology of steroid hormone synthesis. *Endocr. Rev.* **9**, 295–318 (1988).
- Arakane, F. *et al.* Steroidogenic acute regulatory protein (StAR) retains activity in the absence of its mitochondrial targeting sequence: Implications for the mechanism of StAR action. *Proc. Natl Acad. Sci. USA* **93**, 13731–13736 (1996).
- Kallen, C. B. *et al.* Steroidogenic acute regulatory protein (StAR) is a sterol transfer protein. *J. Biol. Chem.* **273**, 26285–26288 (1998).
- Miller, W. L. & Strauss, J. F. III Molecular pathology and mechanism of action of the steroidogenic acute regulatory protein, StAR. *J. Steroid Biochem. Mol. Biol.* **69**, 131–141 (1999).
- Bose, H. S., Whittall, R. M., Baldwin, M. A. & Miller, W. L. The active form of the steroidogenic acute regulatory protein, StAR, appears to be a molten globule. *Proc. Natl Acad. Sci. USA* **96**, 7250–7255 (1999).
- Song, M., Shao, K., Muejeb, A., James, T. L. & Miller, W. L. Molten globule structure and membrane binding of the N-terminal protease-resistant domain (63–193) of the steroidogenic acute regulatory protein (StAR). *Biochem. J.* **356**, 151–158 (2001).
- Christensen, K., Bose, H. S., Harris, F. M., Miller, W. L. & Bell, J. D. Binding of StAR to synthetic membranes suggests an active molten globule. *J. Biol. Chem.* **276**, 17044–17051 (2001).
- Tsujishita, Y. & Hurley, J. H. Structure and lipid transport mechanism of a StAR-related domain. *Nature Struct. Biol.* **7**, 408–414 (2000).
- Harikrishna, J. A., Black, S. M., Szklarz, G. D. & Miller, W. L. Construction and function of fusion enzymes of the human cytochrome P450_{scc} system. *DNA Cell Biol.* **12**, 371–379 (1993).
- Abe, Y. *et al.* Structural basis of presequence recognition by the mitochondrial import receptor Tom20. *Cell* **100**, 551–560 (2000).
- Jefcoate, C. R., Simpson, E. R. & Boyd, G. S. Spectral properties of rat adrenal-mitochondrial cytochrome P-450. *Eur. J. Biochem.* **42**, 539–551 (1974).
- Toaff, M. E., Schleyer, H. & Strauss, J. F. III Metabolism of 25-hydroxycholesterol by rat luteal mitochondria and dispersed cells. *Endocrinology* **111**, 1785–1790 (1982).
- Bose, H. S., Whittall, R. M., Huang, M. C., Baldwin, M. A. & Miller, W. L. N-218 MLN64, a protein with StAR-like steroidogenic activity is folded and cleaved similarly to StAR. *Biochemistry* **39**, 11722–11731 (2000).
- Rapoport, D. & Neupert, W. Biogenesis of Tom40, core component of the Tom complex of mitochondria. *J. Cell. Biol.* **146**, 321–331 (1999).
- Li, J. M. & Shore, G. C. Reversal of the orientation of an integral protein of the mitochondrial outer membrane. *Science* **256**, 1815–1817 (1992).
- Meisinger, C. *et al.* Protein import channel of the outer mitochondrial membrane a highly stable Tom 40-Tom 20 core structure differentially interacts with preproteins, small Tom proteins and import receptors. *Mol. Cell Biol.* **21**, 2337–2348 (2001).

24. Regan, C. I. W. M., Darley-Usmar, V. M. & Lowe, P. N. in *Mitochondria: A Practical Approach* (eds Darley-Usmar, V. M., Rickwood, D. & Wilson, M. T.) 79–112 (IRL Press, Washington DC, 1987).
25. Artemenko, I. P., Zhao, D., Hales, D. B., Hales, K. H. & Jefcoate, C. R. Mitochondrial processing of newly synthesized steroidogenic acute regulatory protein (StAR), but not total StAR, mediates cholesterol transfer to cytochrome P450 side chain cleavage enzyme in adrenal cells. *J. Biol. Chem.* **276**, 46583–46596 (2001).
26. Bose, H. S., Baldwin, M. A. & Miller, W. L. Incorrect folding of steroidogenic acute regulatory protein (StAR) in congenital lipid adrenal hyperplasia. *Biochemistry* **37**, 9768–9775 (1998).
27. Hurt, E. C., Pesold-Hurt, B., Suda, K., Oppliger, W. & Schatz, G. The first twelve amino acids (less than half of the pre-sequence) of an imported mitochondrial protein can direct mouse cytosolic dihydrofolate reductase into the yeast mitochondrial matrix. *EMBO J.* **4**, 2061–2068 (1985).
28. Chung, B., Matteson, K. J., Voutilainen, R., Mohandas, T. K. & Miller, W. L. Human cholesterol side-chain cleavage enzyme, P450_{scc}: cDNA cloning, assignment of the gene to chromosome 15, and expression in the placenta. *Proc. Natl Acad. Sci. USA* **83**, 8962–8966 (1986).
29. Chuck, S. & Lingappa, V. Apolipoprotein B intermediates. *Nature* **356**, 115–116 (1992).
30. Luciano, P. et al. Functional reconstitution of the import of the yeast ADP/ATP carrier mediated by the TIM 10 complex. *EMBO J.* **20**, 4099–4016 (2001).

Acknowledgements

We thank G. C. Shore for the Tom20 antiserum. This work was supported by the National Institutes of Health (H.B., V.R.L. and W.L.M.), the American Heart Association and the Sandler Foundation (V.R.L.), and the UCSF Academic Senate (W.L.M.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and request for materials should be addressed to W.L.M. (e-mail: wmlab@itsa.ucsf.edu).

Neutrophil elastase targets virulence factors of enterobacteria

Yvette Weinrauch*, Doreen Drujan*, Steven D. Shapiro†, Jerrold Weiss‡ & Arturo Zychlinsky*§

* Skirball Institute and Department of Microbiology, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA

† Department of Medicine, Pulmonary and Critical Care, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Departments of Internal Medicine and Microbiology, The Inflammation Program, University of Iowa College of Medicine, Iowa City Veterans' Administration Medical Center, 200 Hawkins Drive, Iowa City, Iowa 52242, USA

§ Max Planck Institute for Infection Biology, Schumannstrasse 21/22, Campus Charité Mitte, Berlin D-10117, Germany

Shigellae cause bacillary dysentery, a bloody form of diarrhoea that affects almost 200 million people and causes nearly 2 million deaths per year¹. *Shigella* invades the colonic mucosa, where it initiates an acute inflammation, rich in neutrophils, that initially contributes to tissue damage and eventually resolves the infection². Neutrophils are phagocytic cells that kill micro-organisms^{3,4} but it is unclear how neutrophils control pathogenic bacteria expressing virulence factors that manipulate host cells. In contrast to other cells, neutrophils prevent the escape of *Shigella* from phagocytic vacuoles in which the bacteria are killed⁵. Here we identify human neutrophil elastase (NE) as a key host defence protein: NE degrades *Shigella* virulence factors at a 1,000-fold lower concentration than that needed to degrade other bacterial proteins. In neutrophils in which NE is inactivated pharmacologically or genetically, *Shigella* escapes from phagosomes, increasing bacterial survival. NE also preferentially cleaves virulence factors of *Salmonella* and *Yersinia*. These findings establish NE as the first neutrophil factor that targets bacterial virulence proteins.

Shigella escapes from the phagosome to the cytoplasm of macrophages and epithelial cells within minutes. In these host cells,

Shigella recruits the cytoskeleton to move intracellularly and inter-cellularly. The genes required for invasion (invasion plasmid antigens *ipaA*, *ipaB*, *ipaC* and *ipaD*) as well as for intracellular movement (*icsA*) are encoded on a virulence plasmid⁶. IpaA, IpaB and IpaC are exported through type III secretion and interact with the host cell. The type III secretion apparatus is composed of at least 30 proteins and is conserved among enteropathogenic bacteria⁷. IcsA, which also interacts with host cell proteins, is secreted through a type V system⁸. Unlike other host cell types, neutrophils prevent *Shigella* from escaping the phagosome, although the bacteria are viable for up to an hour in this compartment⁵. These findings prompted us to hypothesize that neutrophils destroyed the virulence factors of *Shigella* before bacterial viability was affected.

To mimic *in vivo* conditions *in vitro*, we tested different concentrations of a human neutrophil extract enriched in granule proteins (hNEGP) on *Shigella*. *Shigella* was effectively killed at high concentrations of hNEGP, but concentrations below 1% did not affect bacterial viability (Fig. 1a). Sublethal concentrations of hNEGP (up to 0.5% v/v) substantially decreased the levels of type III secreted proteins, IpaA, IpaB and IpaC, in culture supernatants (Fig. 1b). This was surprising because the addition of various other proteins increase type III secretion⁹. The intracellular pool of these proteins was not affected (Fig. 1b). In addition, levels of the outer membrane (relative molecular mass 120,000 (M_r 120K)) (bacterial pellets), and the M_r 90K secreted (culture supernatants) forms of IcsA were also decreased at hNEGP concentrations as low as 0.01% v/v. Sublethal concentrations of hNEGP did not affect outer-membrane protein A (OmpA), maltose-binding protein (MBP) or recombinase A (RecA), which are outer-membrane, periplasmic and cytosolic proteins respectively, confirming that bacterial cell integrity was maintained (Fig. 1c). Interestingly, OmpA and IcsA are both in the outer membrane, indicating that hNEGP specifically targeted virulence proteins.

To test whether neutral serine proteases, which are abundant in neutrophil granules¹⁰, were responsible for the degradation of *Shigella* virulence proteins, we used phenylmethylsulphonyl

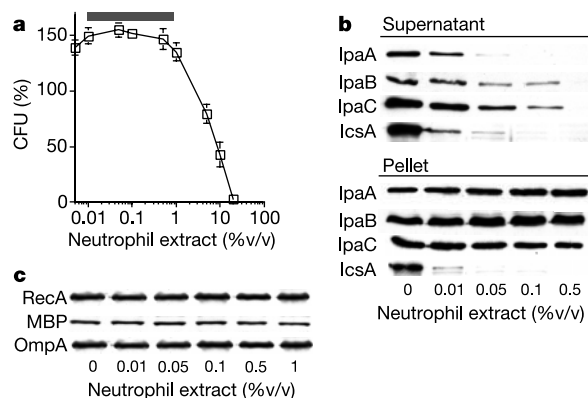


Figure 1 Human neutrophil extract enriched in granule proteins (hNEGP) targets virulence proteins of *Shigella*. **a**, hNEGP kills *Shigella*. The bacterial viability of 10^8 CFU of wild-type *Shigella flexneri* (M90T) after incubation for 30 min at 37 °C decreases at concentrations above 1% v/v hNEGP. The bar indicates the sublethal concentrations of hNEGP that were used in immunoblot analysis shown in **b**. **b**, Sublethal concentrations of hNEGP decrease *Shigella* virulence proteins. Proteins from bacterial pellets and filtered culture supernatants were resolved by SDS-PAGE and analysed by immunoblotting. The type III secreted proteins Ipa A, IpaB and IpaC show a substantial decrease (at 0.05% v/v) in the supernatant but not in the intracellular fraction. The secreted (supernatant) and outer membrane (pellet) forms of IcsA are also degraded at 0.05% v/v. **c**, Bacterial cell integrity is not compromised at sublethal concentrations of hNEGP. Immunoblot analysis of cell pellets, with specific antibodies against OmpA, MBP and RecA, which are respectively outer-membrane, periplasmic and cytosolic marker proteins, show no hNEGP-dependent degradation.

fluoride (PMSF), a group-specific serine protease inhibitor. PMSF blocked the degradation of IpaA, IpaB and IpaC, as well as the membrane-bound and cleaved forms of IcsA (Fig. 2a).

Using specific inhibitors, we identified NE as the protease that degrades *Shigella* effectors. Pretreatment of hNEGP with chemical NE inhibitors (*N*-methoxysuccinyl-Ala-Ala-Pro-Val-chloromethylketone (MeOSuc-AAPV-cmk)¹¹, the trifluoromethyl ketone-based ICI-200355 (ref. 12) and the physiological inhibitor, human secretory leukocyte protease inhibitor (SLPI)¹³, effectively blocked degradation of the *Shigella* proteins (Fig. 2b; only IpaB and the membrane-bound form of IcsA are shown, but IpaA, IpaC and the secreted form of IcsA behaved equivalently). SLPI blocks NE and cathepsin G but not proteinase 3, another neutrophil neutral protease¹³.

Purified human NE, at concentrations as low as 1.2 nM, cleaved *Shigella* virulence proteins significantly (Fig. 2c; only IpaB and IcsA are shown). This concentration mirrors the amount of NE present in sublethal concentrations of hNEGP. Belaouaj *et al.*¹⁴ observed

that high concentrations of NE partly digest OmpA of non-pathogenic *Escherichia coli* (2 μ M, 4 h with 10^6 bacteria). We observed that IcsA was degraded at 1.2 nM NE (30 min incubation with 10^8 bacteria). Under these conditions OmpA was not degraded in either *Shigella* or *E. coli* (data not shown), indicating that NE preferentially cleaved virulence proteins (Fig. 2c). NE also cleaves a broad spectrum of matrix macromolecules, activates some antimicrobial peptides¹⁵ and is implicated in tissue injury¹⁰. In contrast to NE, purified cathepsin G, another prominent neutral serine protease in hNEGP, did not degrade the *Shigella* effectors (Fig. 2c).

NE did not target the type III secretion apparatus itself, which consists of a basal component and a needle that extends beyond the outer membrane^{16,17}. NE did not degrade MxiA or MxiD, respectively the inner and outer membrane needle proteins of *Shigella*, or InvG, a type III component of *Salmonella* (Fig. 2d).

NE is responsible for the degradation of *Shigella* virulence factors in intact neutrophils, confirming the data with hNEGP. IpaA, IpaB and IcsA, but not the cytoplasmic (RecA) or outer membrane

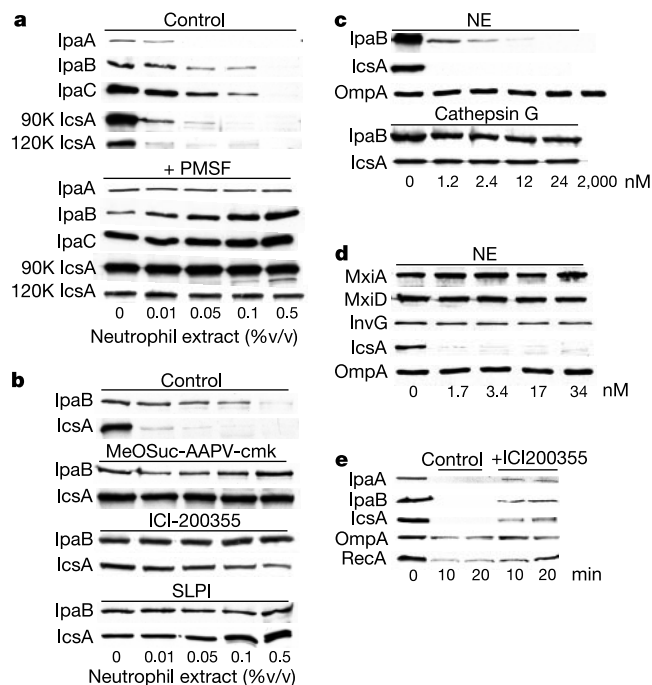


Figure 2 NE degrades *Shigella* virulence proteins. **a**, Treatment of hNEGP with the serine protease inhibitor PMSF (1 mM) blocked the degradation of secreted IpaA, IpaB, IpaC and the outer-membrane (M_r 120K) and secreted (M_r 90K) forms of IcsA. **b**, *Shigella* virulence proteins are degraded by NE in hNEGP. *Shigella* cells (10^8 CFU) were incubated with hNEGP, pretreated with the NE-specific chemical inhibitors MeOSuc-AAPV-cmk or ICI-200355 or the physiological inhibitor SLPI (1 mM, 20 μ M and 750 nM, respectively) for 20 min at 22 °C. Degradation of IpaB and IcsA was blocked by these inhibitors, as shown by immunoblot analysis. **c**, Purified NE, but not cathepsin G, another neutral protease abundant in neutrophil granules, specifically cleaved *Shigella* virulence proteins. *Shigella* cells (10^8 cathepsin FU) were incubated with purified NE or CG at the indicated concentrations. After incubation for 30 min, proteins from culture supernatants and cell pellets were analysed by immunoblot analysis. OmpA was tested in cell pellets as a negative control. **d**, NE does not degrade the type III secretion apparatus. Immunoblots of bacterial pellets treated as in **c** were probed with antibodies against the *Shigella* (MxiA and MxiD) or *Salmonella* (InvG) type III proteins. IcsA and OmpA were the positive and negative controls, respectively, for NE targets in the outer membrane. **e**, *Shigella* virulence factors are degraded by NE in intact neutrophils. Neutrophils (10^6 ml⁻¹), pretreated with ICI-200355 (20 μ M), were infected with wild-type *Shigella* (100 bacteria per neutrophil). At the indicated time points (10 and 20 min), proteins from the filtered culture supernatants (IpaA and IpaB) or the bacterial pellets (IcsA, RecA and OmpA) were analysed by immunoblotting.

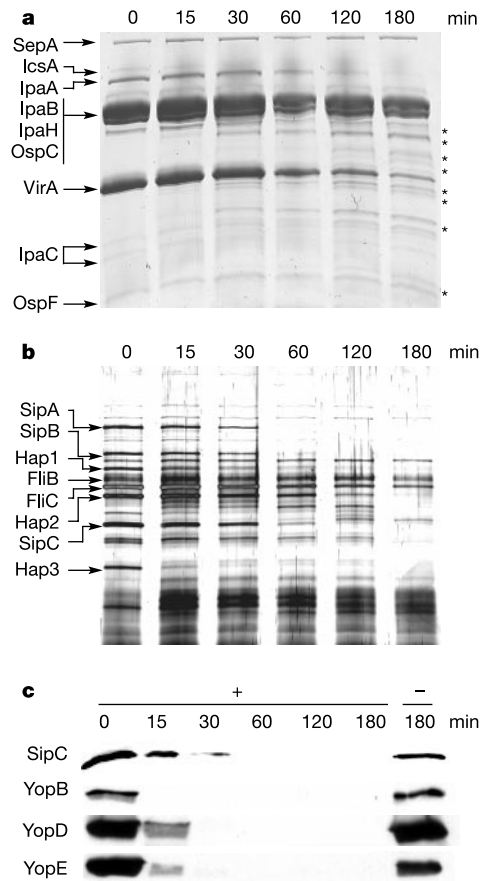


Figure 3 NE degrades secreted virulence proteins of Gram-negative bacterial pathogens. **a**, Selective degradation of virulence proteins in *Shigella* supernatants. Filtered secreted proteins from *Shigella* incubated with 3.4 nM purified NE for the indicated durations. Precipitated proteins were separated by SDS-PAGE and stained with Coomassie blue. Proteins identified by MALDI-TOF mass spectrometry are labelled, asterisks indicate discrete cleavage products. **b**, NE preferentially degrades virulence proteins in *Salmonella*. Secreted proteins from wild-type *S. typhimurium* (strain SL1344) were incubated with purified NE and analysed as described in **a**. Precipitated proteins were resolved by SDS-PAGE and stained with silver nitrate before analysis by MALDI-TOF mass spectrometry. **c**, NE degrades virulence proteins from *S. typhimurium* and *Yersinia enterocolitica*. Secreted proteins from wild-type *S. typhimurium* or *Y. enterocolitica* were incubated with purified NE as described in **a**, and examined by immunoblot analysis with specific antibodies against SipC in *S. typhimurium* and YopB, YopD and YopE in *Y. enterocolitica*. (–) indicates incubation without NE.

(OmpA) markers, were degraded within 10 min of infection of human neutrophils with *Shigella*. This degradation was dependent on NE because the cell-permeable NE inhibitor ICI-200355 blocked it (Fig. 2e).

It is not known how NE recognizes its substrates because the only requirement for cleavage is either an alanine or a valine residue in the P1 position. Purified NE showed specificity by selectively

degrading virulence factors in *Shigella* culture supernatants (Fig. 3a). Identification of the secreted proteins by matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectrometry showed that virulence factors (for example, VirA and the Ipas) were cleaved, but proteins that have not been associated with virulence (SepA and OspF) remained intact (Fig. 3a). NE degradation was not dependent upon the amount of substrate because SepA and OspF were present in higher amounts than, for example, the virulence factors IcsA or IpaC. The repertoire of non-virulent proteins is small because proteins secreted by wild-type *Shigella* consist mostly of virulence factors, whereas non-virulent strains secrete few proteins. The appearance of discrete cleavage products (asterisk in Fig. 3a) suggested that the *Shigella* proteins were folded in the culture supernatants and that only some of the many potential cleavage sites were initially attacked by NE.

NE also targeted proteins secreted by two other Gram-negative pathogens, *Salmonella* and *Yersinia*. The *Salmonella* virulence proteins SipA, SipB, SipC and HAPs, required for flagellar structure (Fig. 3b, c), as well as the *Yersinia* virulence proteins YopB, YopD and YopE, were also degraded by NE (Fig. 3c). Interestingly, FliB and FliC in *Salmonella*, which activate the innate immune receptor TLR5 (ref. 18), were not degraded by NE (Fig. 3b). These results indicate that NE targets virulence proteins from different pathogens. The role of NE in *Salmonella* and *Yersinia* infections, as well as the susceptibility of effectors of other pathogenic bacteria, remains to be determined.

At low multiplicities of infection, *Shigella* is trapped in the phagolysosome of human neutrophils (Fig. 4a)⁵. On the basis of our results, we hypothesized that if NE was inactivated, *Shigella* Ipa proteins would remain functional and the bacteria would escape from the phagosome into the neutrophil cytoplasm. Indeed, when human neutrophils were preincubated with ICI-200355 before infection with wild-type *Shigella*, the bacteria were found free in the cytoplasm (Fig. 4c). In these infections, no *Shigella* cells were found in the cytosol of control neutrophils, whereas 16% of the bacteria were in the cytosol of neutrophils where NE was inhibited. Although *Shigella* is a human-specific pathogen and murine neutrophils lack¹⁹ the components of the non-oxidative arsenal of human neutrophils essential to kill *Shigella*⁵ and other Gram-negative bacteria⁴, we used NE-null mice to test¹⁴ whether, in the absence of NE, neutrophils could contain *Shigella* in the phagosome. Whereas all the *Shigella* cells were contained within the phagosome of murine wild-type neutrophils (Fig. 4d), 12% of the *Shigella* in neutrophils from NE-null mice were clearly found free in the cytoplasm (Fig. 4f). The visibility of the bacterial double membrane in Fig. 4c, f, but not that of the surrounding vacuolar membrane, shows that the *Shigella* cells were free in the cytoplasm. A noninvasive strain of *Shigella* used as a control remained in the phagosome in human neutrophils incubated with ICI-200355 (Fig. 4b) or in murine NE-null neutrophils (Fig. 4e).

Because neutrophils in which NE was inactivated allowed the escape of wild-type *Shigella* into the cytoplasm, we expected a corresponding increase in *Shigella* survival. We found that *Shigella* did indeed survive better both in neutrophils treated with ICI-200355 and in NE-null neutrophils than in controls (Fig. 4g, h). At high multiplicities of infection, *Shigella* is cytotoxic to human²⁰, but not murine, neutrophils. In human neutrophils in which NE was blocked with ICI-200355 and *Shigella* colonized the cytoplasm, cytotoxicity to neutrophils was enhanced (Fig. 4i). As expected, the noninvasive strain was not cytotoxic. Taken together, these data support a prominent role for NE in controlling *Shigella* infections.

The structural basis of the exquisite sensitivity of virulence proteins to NE is not known. Common structural patterns for virulence factors have been proposed^{21,22} and might serve as recognition sites for NE action.

The ability of neutrophils to sequester and kill bacteria is a critical aspect of their defence function. This study demonstrates a

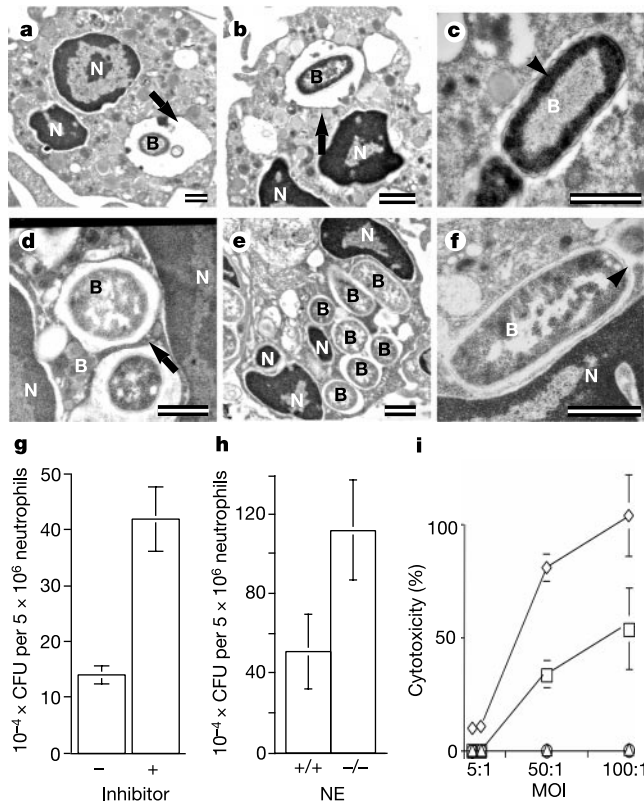


Figure 4 Abrogation of NE permits *Shigella* to escape the phagosome of neutrophils. **a, b, d, e**, Bacteria (labelled B) contained within vacuoles surrounded by vacuolar membranes (arrows) in human neutrophils infected with wild-type *Shigella* (**a**) or a non-virulent strain (**b**), in wild-type murine neutrophils infected with wild-type *Shigella* (**d**) and in murine NE-null neutrophils infected with a non-virulent strain (**e**). **c, f**, Free bacteria in the host cytoplasm in human neutrophils pretreated with the NE elastase inhibitor ICI-200355 (20 μM) before infection with wild-type *Shigella* (**c**) and in murine NE-null neutrophils infected with wild-type *Shigella* (**f**). Arrowheads point to the double membrane characteristic of enterobacteria. Cells were fixed 30 min after infection. N, neutrophil nuclei. Scale bars, 1 μm . **g**, Increased intracellular survival of *Shigella* in human neutrophils in which NE is blocked. Neutrophils, preincubated with (+) or without (-) ICI-200355, were infected with wild-type *Shigella* in duplicate. After 15 min the neutrophils from one sample were washed and cell-associated CFU were determined. The second sample was incubated with gentamicin for an additional 30 min to kill the extracellular bacteria. The values reflect the ratio of intracellular bacteria (samples incubated with gentamicin) to intracellular plus membrane-attached bacteria (samples incubated without gentamicin). **h**, Increased intracellular survival of *Shigella* in murine NE-null neutrophils. Peritoneal neutrophils from null mice and isogenic controls were infected with wild-type *Shigella* as described for **g**. **i**, Increased cytotoxicity of *Shigella* in human neutrophils when NE is blocked. At all multiplicities of infection (MOI) tested, pretreatment of neutrophils with the NE inhibitor ICI-200355 (20 μM) before infection with wild-type *Shigella* (diamonds) resulted in more cell death than did wild-type *Shigella* infection in control neutrophils (squares). Noninvasive *Shigella* control strains caused no cell death, regardless of whether neutrophils were pretreated with ICI-200355 before infection (triangles) or not (circles). Cytotoxicity was determined by the release of cytoplasmic lactate dehydrogenase (LDH) after incubation for 2 h. Data in **g**, **h** and **i** are means and s.d. for triplicate determinations and are representative of a minimum of three experiments with similar results.

previously unrecognized ability of NE to degrade type III secreted and surface-exposed virulence factors of enterobacteria. The ability of NE to destroy these virulence factors rapidly suggests that this enzyme is important in disarming pathogens. □

Methods

Bacterial strains and growth conditions

M90T, an invasive isolate of *S. flexneri* serotype 5, BS176, the noninvasive derivative of M90T and the *Shigella ipaD* mutant²³, which constitutively secretes the Ipa proteins, was grown to exponential phase in tryptic soy broth (TSB), with aeration. *Salmonella typhimurium*, strain SL1344, was grown overnight in Luria-Bertani medium at 37 °C without agitation and *Yersinia enterocolitica* (strain W22703) was grown at 22 °C to a D_{600} of 0.4 in TSB supplemented with 5 mM EGTA and 20 mM $MgCl_2$. The bacteria were centrifuged and the cell pellet was resuspended in nutrient broth supplemented with phosphate-buffered (20 mM, pH 7.4) physiological saline (10^8 colony-forming units (CFU) ml^{-1}) and cultures were shifted to 37 °C for 2 h.

Bactericidal activity

Bactericidal activity of hNEGP prepared as described in ref. 24 was quantified with wild-type strain M90T. In brief, 10^8 bacteria were incubated in a total volume of 1 ml for 30 min at 37 °C with shaking, as described². After aliquots had been removed for determination of CFU, the sample was centrifuged (5 min at 14,000g). Supernatants were passed through a filter (0.2 μm pore size) and recovered proteins were precipitated with methanol/chloroform²⁵.

Protein preparation and immunoblot analysis

Protein from bacterial pellets (2.5×10^7 cell equivalents) and culture supernatants (1 ml) were subjected to SDS–polyacrylamide gel (12.5%) electrophoresis (SDS–PAGE). The protein bands separated by SDS–PAGE were transferred to a nitrocellulose membrane and detected with antibodies specific for IpaA, IpaB, IpaC and IcsA (all gifts from P.J. Sansonetti). Anti-OmpA, anti-MBP and anti-RecA were provided by K. S. Kim, H. Shuman and K. Adzuma, respectively. Alternatively, secreted proteins (30 μg) from culture supernatants from the indicated strains were filtered after separation by centrifugation of bacterial cultures and treated with 3.4 nM NE, after which the reactions were stopped with the addition of PMSF (1 mM) at the indicated times. Proteins were then precipitated as described above, resolved by SDS–PAGE and stained with Coomassie blue or silver nitrate or subjected to immunoblot analysis with SipC, (gift from S. Daefler) or YopB, YopD and YopE antisera (gifts from O. Schneewind). Immunoblotting for type III components was detected in bacterial pellets with antibodies against MxiD (provided by C. Sasakawa), InvG (provided by S. Daefler) and LcrD (provided by O. Schneewind). MxiA was detected with the anti-LcrD antibody, because these proteins are homologous²⁶. The anti-LcrD antibody recognizes a band of the correct relative molecular mass in wild-type *Shigella*, which is absent from the non-virulent strain BS176 because it lacks the plasmid where MxiA is encoded.

Enzymatic assays

NE was quantified in hNEGP by using the substrate *N*-methoxysuccinyl-Ala-Ala-Pro-Val-*p*-nitroanilide in accordance with manufacturer's protocols (Elastin Products Company). NE activity in hNEGP was completely inhibited by PMSF (1 mM, 22 °C, 20 min), MeOSuc-AAPV-cmk (Sigma), ICI-200355 (ref. 27) (provided by AstraZeneca Pharmaceuticals) or SLPI (R&D Systems). Cathepsin G (Elastin Products Company) was assayed in accordance with the manufacturer's instructions. The NE activity of intact neutrophils was determined as described previously²⁷. NE did not affect bacterial viability even at 2,000 nM (data not shown).

Isolation of neutrophils

Human neutrophils (>95% pure) were isolated from the peripheral blood of healthy donors by using dextran-Ficoll²⁸, and resuspended in complete culture medium (CM, consisting of Dulbecco's modified Eagle's medium plus 10% fetal bovine serum, before stimulation. Murine thyoglycollate-elicited peritoneal neutrophils (>80% pure) were isolated from NE-null or isogenic controls, resuspended at $5 \times 10^6 ml^{-1}$ and left to adhere to plastic plates in CM. Human neutrophils were resuspended at $5 \times 10^6 ml^{-1}$ and left to adhere to plastic plates in CM containing the biologically relevant²⁹ activator interleukin-8 (10 nM; R&D Systems). After incubation for 30 min at 37 °C, the medium was replaced with serum-free culture medium (SFM), with or without ICI-200355 (ref. 27) (20 μM), for 30 min before infection. ICI-200355 is not cytotoxic (as measured by lactate dehydrogenase release (Promega)) and does not affect the ingestion or killing of non-virulent bacteria²⁷, but it blocks 80% of NE activity²⁷. Human or murine neutrophils were infected with M90T or BS176 in SFM, centrifuged at 700g for 10 min and incubated at 37 °C and 5% CO_2 . At 30 min after infection, neutrophils were washed, fixed and processed for transmission electron microscopy with the use of standard methods. For the measurement of bacterial viability, duplicate samples of neutrophils were infected (10 bacteria per neutrophil) with wild-type *Shigella*. One sample was washed and plated 15 min after incubation to determine the number of cell-associated bacteria and the second sample after incubation for 30 min with gentamicin (100 $\mu g ml^{-1}$) to determine the number of intracellular bacteria. Infections were performed in SFM because *Shigella* invasion is independent of serum opsonins³⁰. To test the degradation of virulence factors in cell-culture experiments, interleukin-8-stimulated neutrophils (10^6) with or without ICI-200355 were infected with wild-type *Shigella* (10^8) in a total volume of 1 ml. At the indicated times, the proteins in filtered culture supernatants were precipitated as described

above. Cell pellets (neutrophils and bacterial) or supernatant proteins were solubilized in SDS sample buffer and processed for immunoblot analysis.

Received 27 December 2001; accepted 6 February 2002.

- Kotloff, K. L. *et al.* Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. World Health Org.* **77**, 651–666 (1999).
- Perdomo, O. J. *et al.* Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. *J. Exp. Med.* **180**, 1307–1319 (1994).
- Klebanoff, S. J. In *Inflammation: Basic Principles and Clinical Correlates* (eds Gallin, J. I. & Snyderman, R.) 721–768 (Lippincott Williams & Wilkins, Philadelphia, 1999).
- Elsbach, P., Weiss, J. & Levy, O. In *Inflammation: Basic Principles and Clinical Correlates* (eds Gallin, J. I. & Snyderman, R.) 801–817 (Lippincott Williams & Wilkins, Philadelphia, 1999).
- Mandic-Mulec, I., Weiss, J. & Zychlinsky, A. *Shigella flexneri* is trapped in polymorphonuclear leukocyte vacuoles and efficiently killed. *Infect. Immun.* **65**, 110–115 (1997).
- Hale, T. L. Genetic basis of virulence in *Shigella* species. *Microbiol. Rev.* **55**, 206–224 (1991).
- Cornelis, G. R. & Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**, 735–774 (2000).
- Suzuki, T., Lett, M. C. & Sasakawa, C. Extracellular transport of VirG protein in *Shigella*. *J. Biol. Chem.* **270**, 30874–30880 (1995).
- Bahrani, F. K., Sansonetti, P. J. & Parsot, C. Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect. Immun.* **65**, 4005–4010 (1997).
- Owen, C. A. & Campbell, E. J. The cell biology of leukocyte-mediated proteolysis. *J. Leukoc. Biol.* **65**, 137–150 (1999).
- Campbell, E. J., Silverman, E. K. & Campbell, M. A. Elastase and cathepsin G of human monocytes. Quantification of cellular content, release in response to stimuli, and heterogeneity in elastase-mediated proteolytic activity. *J. Immunol.* **143**, 2961–2968 (1989).
- Veale, C. A. *et al.* Orally active trifluoromethyl ketone inhibitors of human leukocyte elastase. *J. Med. Chem.* **40**, 3173–3181 (1997).
- Wright, C. D., Kennedy, J. A., Zitnik, R. J. & Kashem, M. A. Inhibition of murine neutrophil serine proteinases by human and murine secretory leukocyte protease inhibitor. *Biochem. Biophys. Res. Commun.* **254**, 614–617 (1999).
- Belaaouaj, A., Kim, K. S. & Shapiro, S. D. Degradation of outer membrane protein A in *Escherichia coli* killing by neutrophil elastase. *Science* **289**, 1185–1188 (2000).
- Panyutich, A., Shi, J., Boutz, P. L., Zhao, C. & Ganz, T. Porcine polymorphonuclear leukocytes generate extracellular microbicidal activity by elastase-mediated activation of secreted propeptidases. *Infect. Immun.* **65**, 978–985 (1997).
- Tamano, K. *et al.* Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. *EMBO J.* **19**, 3876–3887 (2000).
- Blocker, A. *et al.* Structure and composition of the *Shigella flexneri* 'needle complex', a part of its type III secretin. *Mol. Microbiol.* **39**, 652–663 (2001).
- Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–1103 (2001).
- Eisenhauer, P. B., Harwig, S. S. & Lehrer, R. I. Cryptidins: Antimicrobial defensins of the murine small intestine. *Infect. Immun.* **60**, 3556–3565 (1992).
- Francois, M., Le Cabec, V., Dupont, M. A., Sansonetti, P. J. & Maridonneau-Parini, I. Induction of necrosis in human neutrophils by *Shigella flexneri* requires type III secretion, IpaB and IpaC invasins, and actin polymerization. *Infect. Immun.* **68**, 1289–1296 (2000).
- Pallen, M. J., Dougan, G. & Frankel, G. Coiled-coil domains in proteins secreted by type III secretion systems. *Mol. Microbiol.* **25**, 423–425 (1997).
- Miao, E. A. *et al.* *Salmonella typhimurium* leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. *Mol. Microbiol.* **34**, 850–864 (1999).
- Menard, R., Sansonetti, P. J. & Parsot, C. Nonpolar mutagenesis of the *ipa* genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J. Bacteriol.* **175**, 5899–5906 (1993).
- Weiss, J., Elsbach, P., Olsson, I. & Odeberg, H. Purification and characterization of a potent bactericidal and membrane active protein from the granules of human polymorphonuclear leukocytes. *J. Biol. Chem.* **253**, 2664–2672 (1978).
- Lee, V. T. & Schneewind, O. Type III machines of pathogenic *Yersinia* secrete virulence factors into the extracellular milieu. *Mol. Microbiol.* **31**, 1619–1629 (1999).
- Ginocchio, C. C. & Galan, J. E. Functional conservation among members of the *Salmonella typhimurium* InvA family of proteins. *Infect. Immun.* **63**, 729–732 (1995).
- Huang, Y. I. *et al.* Effect of trifluoromethyl ketone-based elastase inhibitors on neutrophil function in vitro. *J. Leukoc. Biol.* **64**, 322–330 (1998).
- Weiss, J., Kao, L., Victor, M. & Elsbach, P. Oxygen-independent intracellular and oxygen-dependent extracellular killing of *Escherichia coli* S15 by human polymorphonuclear leukocytes. *J. Clin. Invest.* **76**, 206–212 (1985).
- Sansonetti, P. J., Arondel, J., Huerre, M., Harada, A. & Matsushima, K. Interleukin-8 controls bacterial trans epithelial translocation at the cost of epithelial destruction in experimental shigellosis. *Infect. Immun.* **67**, 1471–1480 (1999).
- Gbarah, A. *et al.* *Shigella flexneri* transformants expressing type 1 (mannose-specific) fimbriae bind to, activate, and are killed by phagocytic cells. *Infect. Immun.* **61**, 1687–1693 (1993).

Acknowledgements

We thank D. Weiss and C. Scharff for critical reading of the manuscript, T. Nubert and D. Wu for performing MS analysis, and staff of the Rockefeller University EM facility. A.Z. is a recipient of the Irma T. Hirsch Trust Career Scientist Award. Financial support from the National Institutes of Health is acknowledged.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.Z. (e-mail: zychlinsky@mpiib-berlin.mpg.de).

HIV preferentially infects HIV-specific CD4⁺ T cells

Daniel C. Douek^{*†‡}, Jason M. Brenchley^{*‡}, Michael R. Betts^{*}, David R. Ambrozak^{*}, Brenna J. Hill^{*}, Yukari Okamoto^{*}, Joseph P. Casazza[§], Janaki Kuruppu^{*}, Kevin Kunstman^{||}, Steven Wolinsky^{||}, Zvi Grossman[¶], Mark Dybul[#], Annette Oxenius[☆], David A. Price[☆], Mark Connors[#] & Richard A. Koup^{*}

^{*} Vaccine Research Center and [#] Laboratory of Immunoregulation, NIAID, NIH, and [†] Department of Experimental Transplantation and Immunology, Medicine Branch, NCI, NIH, Maryland 20892, USA

[§] Department of Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

^{||} Department of Infectious Diseases, Northwestern University Medical School, Chicago, Illinois 60611, USA

[¶] Laboratory of Immunology, NIAID, NIH, Maryland, USA, and Department of Physiology and Pharmacology, Tel Aviv University, Tel Aviv 69978, Israel

[☆] Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

[‡] These authors contributed equally to this work

HIV infection is associated with the progressive loss of CD4⁺ T cells through their destruction or decreased production^{1,2}. A central, yet unresolved issue of HIV disease is the mechanism for this loss, and in particular whether HIV-specific CD4⁺ T cells are preferentially affected^{3–5}. Here we show that HIV-specific memory CD4⁺ T cells in infected individuals contain more HIV viral DNA than other memory CD4⁺ T cells, at all stages of HIV disease. Additionally, following viral rebound during interruption of antiretroviral therapy, the frequency of HIV viral DNA in the HIV-specific pool of memory CD4⁺ T cells increases to a greater extent than in memory CD4⁺ T cells of other specificities. These findings show that HIV-specific CD4⁺ T cells are preferentially infected by HIV *in vivo*. This provides a potential mechanism to explain the loss of HIV-specific CD4⁺ T-cell responses, and consequently the loss of immunological control of HIV replication⁶. Furthermore, the phenomenon of HIV specifically infecting the very cells that respond to it adds a cautionary note to the practice of structured therapy interruption.

In a cross-section of 12 HIV-infected individuals (Table 1), we sorted by flow cytometry HIV-specific and cytomegalovirus (CMV)-specific memory CD4⁺ T cells by antigen-induced inter-

feron- γ (IFN- γ) production; we also sorted unstimulated bulk CD4⁺ CD45RO⁺ memory T cells. Using sensitive quantitative polymerase chain reaction (qPCR) to measure viral HIV DNA as a quantitative record of infection history, we found significantly more viral DNA in the HIV-specific CD4⁺ T cells than in CD45RO⁺ CD4⁺ T cells ($P = 0.0001$) or CMV-specific CD4⁺ T cells ($P = 0.0039$) (Fig. 1a). There was no significant difference between CMV-specific and CD45RO⁺ CD4⁺ T cells ($P = 0.5$). Furthermore, the abundance of viral DNA in HIV-specific CD4⁺ T cells correlated with that in memory T cells ($r = 0.97$, $P < 0.0001$), and the ratio between the two varied only between 1.5 and 5. Although the qPCR does not distinguish integrated from unintegrated HIV DNA, it is a record of infection history. Thus in every subject, whether acutely or chronically infected, treated or untreated, HIV-specific CD4⁺ memory T cells had been infected at a 2.1–5.3 times (mean, 3.7) higher frequency than CD4⁺ memory T cells of other specificities (Fig. 1a).

Despite higher frequencies of infection in HIV-specific CD4⁺ memory T cells, these cells comprise a small proportion of all CD4⁺ memory T cells⁷ (Table 1). We calculated that infected HIV-specific CD4⁺ memory T cells constituted only between 1.3% to 10.4% (mean, 5.1%) of all infected CD4⁺ memory T cells in blood (Fig. 1b), which was nevertheless more than the contribution from CMV-specific CD4⁺ memory T cells. We note that the contribution of infected HIV-specific CD4⁺ T cells to the total infected T-cell pool, as determined above, is likely to be a minimal estimate, as not all may produce IFN- γ or even be fully activated, and the majority may be located in lymphoid tissue rather than blood. Therefore, these data show that HIV-specific CD4⁺ T cells contribute substantially to the total cell-associated HIV load.

The events occurring during acute HIV infection may constitute one mechanism for the higher frequency of infection of HIV-specific memory CD4⁺ T cells. At the time of initial infection, HIV-specific CD4⁺ T cells would have a naive phenotype, in contrast to T cells specific for common viruses such as CMV, which would have a memory phenotype. Upon encountering their cognate HIV antigens, naive HIV-specific CD4⁺ T cells would be susceptible to infection as they expand and differentiate into effector T cells, leading to productive as well as latent infection⁸. Indeed, it has been shown that during acute infection there may be partial loss of HIV-specific CD4⁺ T-cell responses^{3,5,9}. To explore this mechanism we assessed the effect of successive rounds of cell division, and acquisition of cytokine-secreting effector function, on HIV infection of naive and memory CD4⁺ T cells which had been activated with staphylococcal enterotoxin B (SEB) *in vitro*. After

Table 1 Clinical details of subjects and frequency of HIV-specific CD4⁺ T-cell responses as a percentage of total CD4⁺ T cells.

ID	Approx. time infected (years)	ART*	Viral load (Log ₁₀ RNA copies ml ⁻¹)	CD4 count (cells μ l ⁻¹)	HIV-specific CD4 ⁺ cells (%)
1	6	+	2.3	700	0.12
2	11	—	3.0	800	0.64
3	14	+	4.0	100	1.18
4	Acute	—	3.9	1,060	1.08
5	3	—	3.0	590	0.55
6	5	+	<1.7	300	0.43
7	10	+	3.5	460	0.94
8	Acute	—	5.6	610	0.99
9	15	—	4.0	560	1.17
10	Acute	—	4.8	170	1.58
11	3	+	4.2	440	1.18
12	6	—	3.9	480	1.06
13	10	+	<1.7	720	0.3
13	10.1	—	4.7	500	1.2
14	15	+	<1.7	750	2.94
14	15.1	—	4.5	650	5.11
15	0.3	+	<1.7	700	0.54
15	0.8	—	4.9	430	1.18
16	12	+	1.9	1,190	0.42
16	12.1	—	3.9	1,020	0.39

Subjects 13–16 underwent STI.

* ART, antiretroviral therapy: treated (+), untreated (—).

infection with X4-tropic virus NL4-3 (Fig. 2a and c), or R5-tropic virus JR-CSF (Fig. 2b and d), we assessed intracellular p24 production over the next 24 h at each round of division using the carboxyfluorescein diacetate succinimidyl ester (CFSE) dye as a marker of proliferation. With each round of division, there were successively more HIV-infected cells. In addition, the percentage of infected activated naive CD4⁺ T cells was considerably higher than that of infected activated memory CD4⁺ T cells.

After activation, naive CD4⁺ T cells begin to secrete effector cytokines, such as IFN- γ , after approximately three rounds of division¹⁰. We found that p24 expression occurred in both IFN- γ ⁻ and IFN- γ ⁺ CD4⁺ T cells, but was more frequent in the IFN- γ ⁻ subset, particularly in the X4-virus infection (Fig. 2e and f for NL4-3 or JR-CSF, respectively). CXCR4 and CCR5 co-receptor levels were similar in both IFN- γ ⁻ and IFN- γ ⁺ populations (data not shown). We confirmed this observation in a separate experiment using qPCR to measure CD4⁺ T-cell-associated viral DNA at each round of division, within X4-virus-infected activated naive CD4⁺ T cells that had been sorted by fluorescence-activated cell sorting (FACS, Fig. 2i). The frequency of viral DNA increased with cell division. However, at the third round of division we found that the IFN- γ ⁺ cells contained nearly four times less viral DNA than IFN- γ ⁻ cells. A lower inoculum of virus was used in this experiment than in the infections where p24 production was measured by flow cytometry. To confirm the internal consistency of our p24 staining and qPCR data, we sorted by FACS known numbers of p24^{hi} cells into individual wells for qPCR: we found a 1:1 relationship between the number of p24^{hi} cells and copies of Gag DNA per well, at cell

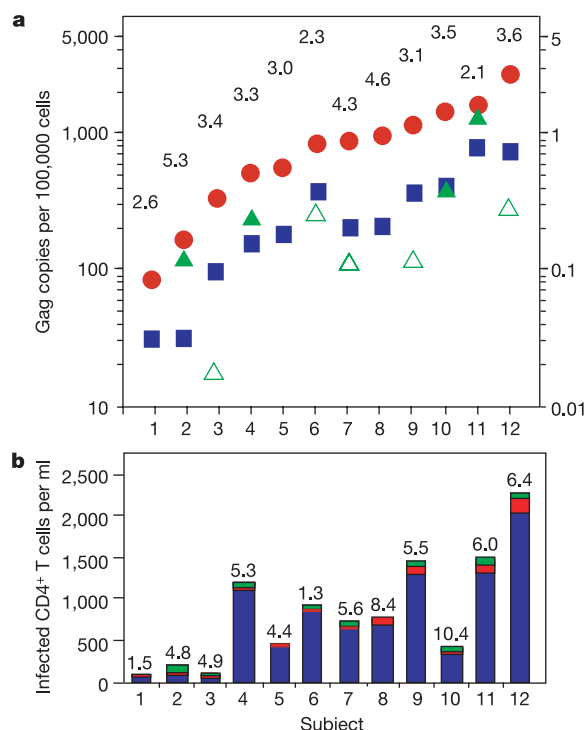


Figure 1 HIV viral DNA content of memory CD4⁺ T cells in infected individuals. **a**, Viral DNA copies were determined by qPCR in HIV-specific (red circles), CMV-specific (green triangles) and CD45RO⁺ memory (blue squares) CD4⁺ T cells. Open symbols denote virus below level of detection (one copy) for number of cells in the PCR, and represent the maximum possible frequency of viral DNA. Not all subjects responded to CMV. Above each point is the ratio of frequency of infection between HIV-specific and CD45RO⁺ CD4⁺ T cells. Also shown is the percentage of infected cells, assuming one viral DNA copy per cell. **b**, The number of infected HIV-specific (red), CMV-specific (green) and CD45RO⁺ memory of other specificities (blue) CD4⁺ T cells per ml blood. Above each bar is the percentage of all HIV-infected memory CD4⁺ T cells that are HIV-specific.

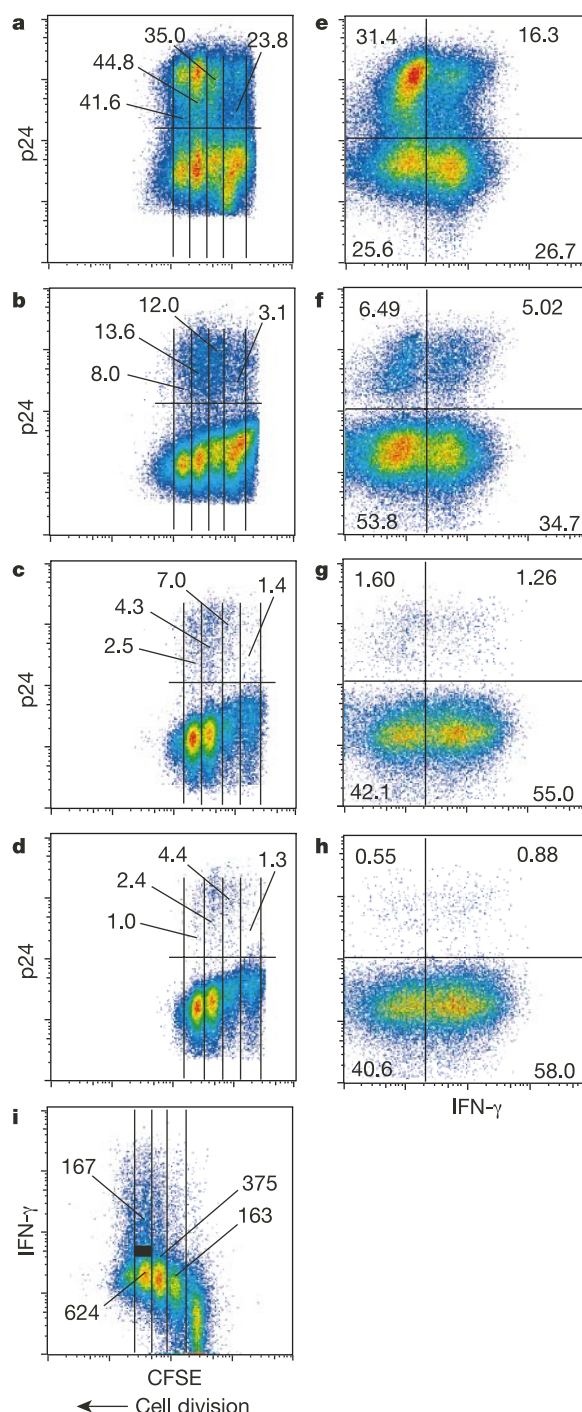


Figure 2 Effects of cell division and interferon- γ secretion on HIV infection of CD4⁺ T cells that had been stimulated by staphylococcal enterotoxin B. **a–d**, Each round of cell division is defined by a twofold decrease in carboxyfluorescein diacetate succinimidyl ester (CFSE) fluorescence, and shown bordered by vertical lines. The percentage of p24^{hi} cells at each division is shown (m.o.i., 0.5). Cells were gated on CD3 and CD4. Uninfected controls (not shown) gave 0.00% p24^{hi} cells at each cell division. Unstimulated cells (not shown) have the same CFSE fluorescence as undivided cells. **a**, Activated naive T cells infected with NL4-3. **b**, Activated naive T cells infected with JR-CSF. **c**, Activated memory T cells infected with NL4-3. **d**, Activated memory T cells infected with JR-CSF. **e–h**, Percentage of p24 production in cells secreting and not secreting IFN- γ (m.o.i., 0.5). **e**, Activated naive T cells infected with NL4-3. **f**, Activated naive T cells infected with JR-CSF. **g**, Activated memory T cells infected with NL4-3. **h**, Activated memory T cells infected with JR-CSF. **i**, Activated naive T cells infected with NL4-3 (m.o.i., 0.05) showing IFN- γ production and cell division. Copies of viral DNA per 100,000 cells determined by qPCR in each population sorted by FACS are shown.

numbers ranging from 100 cells per well down to a single cell per well (data not shown).

Thus HIV-specific naive CD4⁺ T cells, which become activated on first encounter with HIV antigens *in vivo* and are in prolonged close proximity to HIV-containing dendritic cells, would become highly susceptible to HIV infection as they undergo several rounds of division during their initial expansion and differentiation into effector T cells. This could lead to preferential infection of HIV-specific CD4⁺ T cells during acute infection. Unlike naive T cells, activated memory CD4⁺ T cells secrete IFN- γ before division¹⁰. Thus, although the vast majority of memory CD4 T cells infected *in vivo* should harbour an R5 virus, the observation that X4-tropic HIV replicated predominantly in the IFN- γ ⁻ population in activated naive T cells could cast doubt on the appropriateness of measuring viral DNA load in IFN- γ ⁺ memory T cells from HIV-infected individuals (Fig. 1). However, when we infected activated memory CD4⁺ T cells *in vitro*, p24 production was present in both IFN- γ ⁺ and IFN- γ ⁻ subsets (Fig. 2g and h). Therefore it was valid to measure viral DNA in IFN- γ ⁺ memory T cells as a record of their infection history.

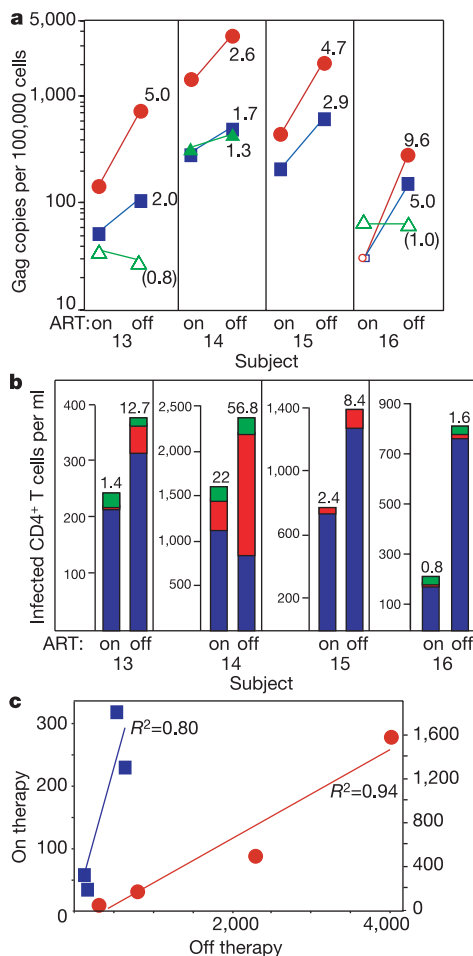


Figure 3 Viral DNA content of memory CD4⁺ T cells in infected individuals on and off antiretroviral therapy, ART. **a**, Viral DNA in HIV-specific (red circles), CMV-specific (green triangles) and CD45RO⁺ memory (blue squares) CD4⁺ T cells. Open symbols denote virus below level of detection, and represent the maximum possible frequency. The number of times increase in viral DNA is shown. **b**, Number of infected HIV-specific (red), CMV-specific (green) and CD45RO⁺ memory of other specificities (blue) CD4⁺ T cells per ml blood. Above each bar is the percentage of all HIV-infected memory CD4⁺ T cells that are HIV-specific. **c**, Relationship between viral DNA frequencies (copies per 100,000 cells) on and off therapy in HIV-specific (red circles, right-hand y axis) and in CD45RO⁺ (blue squares, left-hand y axis) CD4⁺ T cells.

HIV-specific CD4⁺ T-cell responses are often maintained following acute infection and late into chronic infection^{6,7,11–13}, and may be propagated in peripheral lymphoid tissue, where viral replication and T-cell activation is prolific^{14–18}. Local viral replication among responding HIV-specific CD4⁺ T cells may constitute a mechanism to maintain a higher frequency of infection of HIV-specific memory CD4⁺ T cells in chronic infection^{19,20}. Although memory T cells are more susceptible to the cytopathic effects of HIV than naive T cells after activation²¹, clearly not all responsive cells succumb as a result^{6,7,11–13}. Thus, viral DNA present in HIV-specific CD4⁺ T cells could have been established largely during the naive to effector T-cell transition around the time of acute infection (representing a distant archive of infected cells^{22,23}), or could be continually renewed by recurrent infection of established HIV-specific memory CD4⁺ T cells. To address this question, we quantified viral DNA associated with CD4⁺ memory T cells in four individuals who had detectable HIV-specific CD4⁺ T-cell responses while on antiretroviral therapy (ART) and experienced rapid viral rebound during structured therapy interruption (STI; Table 1). In three subjects there was an increase in the frequency of HIV-specific CD4⁺ T cells during STI, confirming previous reports²⁴. The viral DNA frequency increased in both HIV-specific and CD45RO⁺ CD4⁺ memory T cells during STI; however, the increase measured in the HIV-specific CD4⁺ T cells was significantly greater ($P = 0.048$) (Fig. 3a). Viral DNA was detected in CMV-specific CD4⁺ T cells in only one subject, but the increase off therapy was lower than that in HIV-specific and memory CD4⁺ T cells. Additionally, off therapy, the infected HIV-specific CD4⁺ T cells made up a greater fraction of all infected CD4⁺ memory T cells in blood than when on ART (Fig. 3b).

Most notably, when subject no. 14 stopped therapy, more than half of the infected CD4⁺ memory T cells were HIV-specific, suggesting that the HIV-specific CD4⁺ T-cell response *per se* may provide ‘fuel’ for viral spreading. Furthermore, although based on only four subjects, the notable proportionality between the frequencies of virus both in HIV-specific CD4⁺ T cells and in other memory T cells before and after STI (Fig. 3c) suggests that resting, latently infected memory CD4⁺ T cells contribute directly to viral replication^{19,20}. Collectively, these data indicate that HIV-specific CD4⁺ T cells are preferentially infected during viral recrudescence *in vivo*, and suggest that the pool of infected HIV-specific CD4⁺ T cells is continuously replenished. The numbers of infected HIV-specific CD4⁺ T cells sorted in these experiments were much too low to perform accurate molecular analysis of viral RNA species. However, we were able to sort live HIV-specific CD4⁺ T cells from subject no. 7, and detect p24 production after *in vitro* culture, confirming that the virus was indeed replication-competent (data not shown).

Our results provide direct evidence that HIV preferentially infects HIV-specific CD4⁺ T cells *in vivo*. Simply by virtue of their specificity, HIV-specific CD4⁺ T cells are more likely to be in prolonged close proximity to actively replicating HIV—their cognate antigen—in the lymph node^{14–18}. The recruitment of HIV-specific T cells into infected lymphoid sites, enhanced by inflammatory chemoattractants, may affect the rate of viral clearance²⁵ but may also provide cellular substrates for viral replication. During acute infection *in vivo*, rapidly proliferating HIV-specific CD4⁺ T cells, in transition from naive to full effector phenotype, are highly susceptible to HIV infection. Furthermore, although STI has been proposed as a method to boost HIV-specific immunity, to allow repopulation with wild-type virus, and to reduce toxicity from treatment^{26–30}, our data suggest that this may occur at the price of enhancing HIV replication by expanding, rather than controlling, critical pools of infected memory T cells, in particular the pool containing HIV-specific CD4⁺ T cells. Regardless of their preferential infection, the majority of demonstratively antigen-responsive HIV-specific CD4⁺ T cells remain virus-free at any time.

Collectively, these results emphasize the local nature of HIV replication^{19,20}.

These results also provide an empirical basis for a mechanism whereby progressive loss of HIV-specific CD4⁺ T-cell responses may begin during acute infection, and continue throughout its course^{3,6,31}, eventually leading to uncontrolled viral replication and AIDS. Whether such progressive loss indeed occurs has yet to be fully established. In any event, this preferential but low-frequency infection of HIV-specific and other activated T cells may represent an evolutionary adaptation that enables persistent infection in an otherwise immunocompetent host³², and prolonged host–parasite coexistence. □

Methods

Subjects

Subjects defined as acute HIV infection were men with recent exposure, and a negative or evolving HIV-specific antibody response in the presence of viraemia at presentation. Treated subjects were on continuous therapy. All subjects were enrolled in various studies at the University of Texas Southwestern Medical Center and the National Institutes of Health, USA. IRB approval was obtained from each institution. Subject ID was assigned simply according to their order in Fig. 1.

T-cell infection and staining

Naive T cells (depleted of CD45RO⁺ and CD57⁺ cells with magnetic beads to 99% purity), or memory T cells (CD45RA-depleted) from the same healthy subjects were stained with CFSE (0.25 µM, Molecular Probes). Cells were washed and then stimulated with SEB (10 µg ml⁻¹, Sigma) for 48 h at 37 °C. Peripheral blood mononuclear cells (PBMC) were pelleted, infected with NL4-3 or JR-CSF at multiplicity of infection (m.o.i.) of 0.5 for p24 staining or 0.05 for qPCR, and incubated an additional 24 h, followed by 3 h stimulation with phorbol myristyl acetate (PMA; 25 ng ml⁻¹, Sigma), ionomycin (1 µg ml⁻¹, Sigma) and brefeldin A (10 µg ml⁻¹, Sigma). Cells were stained with conjugated antibodies to CD4, then permeabilized (Becton Dickinson), and stained with conjugated antibodies to p24 (Coulter) and IFN-γ (Becton Dickinson). Cells were analysed with a FACS Calibur using FlowJo software.

T-cell stimulation and sorting

Subjects were screened for HIV- and CMV-specific CD4⁺ T-cell responses by intracellular cytokine staining using pools of 15-mer peptides overlapping by 11 residues corresponding to all HIV proteins, or whole CMV as previously described⁷. Up to 2 × 10⁸ PBMCs from subjects were incubated with all relevant peptide pools (Gag, Pol, Env and Nef) to which they responded with frequencies greater than 0.3% to any one or more combined HIV protein, then stained with antibodies to CD3, CD4, CD69 and IFN-γ (Becton Dickinson) and fixed⁷. Antigen-specific cells were sorted with a FACS Vantage SE/ DiVa by tightly gating on CD3⁺ CD4⁺ CD69⁺ IFN-γ⁺ cells outside the range of background staining (anti-CD28/CD49d alone), which was always less than 0.05%. 500–20,000 antigen-specific cells were sorted depending on the original amount of PBMCs. Unstimulated CD3⁺ CD4⁺ CD45RO⁺ 'memory' cells were also stained and sorted for each subject. IFN-γ staining for live cell FACS was done by surface capture (Miltenyi Biotec) according to the manufacturer's instructions. Cells were cultured for 2 weeks with phytohaemagglutinin-stimulated PBMCs and p24 was measured by ultrasensitive enzyme-linked immunosorbent assay (ELISA).

Viral DNA and plasma viral load

HIV DNA was quantified by qPCR with an ABI7700 (Perkin-Elmer). HIV Gag primers and probe were designed against the Los Alamos HIV database. Conserved primers and probe sequences were trimmed to optimal sequences that matched >98% of all sequences in the database at 100% identity. Gag primer position and sequence were 795gagF: ggtgcgagcgtcagtaataag, 911gagR: agctccctgctgccata, and probe was 841gagP: FAM-aaaattcgttaagccaggggaaagaa-QSY7 (MegaBases). To quantify cell number in each reaction, qPCR was performed simultaneously for albumin gene copy number. Albumin primer/probe sequences were AlbF: tgcatgagaaacgcagtaaa, AlbR: atgtgcctgcttaccacaa, and AlbP: FAM-tgacagagtcacaaatgctgcacagaa-QSY7. Sorted T cells were lysed in 30 µl 200 µg ml⁻¹ proteinase K (Boehringer). qPCR was performed on 5 µl cell lysate for 45 cycles using Platinum Taq (Invitrogen). Standards were constructed for absolute quantification of Gag and albumin copy number, and were validated with sequential dilutions of 8E5 and Ach2 cell lysates which contain one copy of Gag per cell. The Gag qPCR was sensitive to one copy per reaction: Ach2 lysate was diluted to 5 Gag copies which were distributed between 10 qPCR reactions, 5 of which contained one copy and 5 were negative. Duplicate reactions were run and template copies calculated by the ABI7700 software. Plasma virus was quantified using the Amplicor RT-PCR kit (Roche Diagnostic Systems). Comparisons in viral DNA loads were assessed by the Wilcoxon matched pairs test, and correlations by the Spearman's rank test, using Prism 3 software.

Received 5 November 2001; accepted 25 February 2002.

- Rowland-Jones, S. HIV infection: where have all the T cells gone? *Lancet* **354**, 5–7 (1999).
- McCune, J. M. The dynamics of CD4⁺ T-cell depletion in HIV disease. *Nature* **410**, 974–979 (2001).
- Rosenberg, E. *et al.* Vigorous HIV-1-specific CD4⁺ T cell responses associated with control of viremia. *Science* **278**, 1447–1450 (1997).

- Rosenberg, E. S. *et al.* Immune control of HIV-1 after early treatment of acute infection. *Nature* **407**, 523–526 (2000).
- Oxenius, A. *et al.* Early highly active antiretroviral therapy for acute HIV-1 infection preserves immune function of CD8⁺ and CD4⁺ T lymphocytes. *Proc. Natl Acad. Sci. USA* **97**, 3382–3387 (2000).
- Pitcher, C. J. *et al.* HIV-1-specific CD4⁺ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression. *Nature Med.* **5**, 518–525 (1999).
- Betts, M. R. *et al.* Analysis of total human immunodeficiency virus (HIV)-specific CD4 and CD8 T-cell responses: relationship to viral load in untreated HIV infection. *J. Virol.* **75**, 11983–11991 (2001).
- Eckstein, D. A. *et al.* HIV-1 actively replicates in naive CD4(+) T cells residing within human lymphoid tissues. *Immunity* **15**, 671–682 (2001).
- Altfield, M. *et al.* Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. *J. Exp. Med.* **193**, 169–180 (2001).
- Bird, J. J. *et al.* Helper T cell differentiation is controlled by the cell cycle. *Immunity* **9**, 229–237 (1998).
- Kalams, S. A. *et al.* Association between virus-specific cytotoxic T-lymphocyte and helper responses in human immunodeficiency virus type 1 infection. *J. Virol.* **73**, 6715–6720 (1999).
- Wilson, J. D. *et al.* Loss of CD4⁺ T cell proliferative ability but not loss of human immunodeficiency virus type 1 specificity equates with progression to disease. *J. Infect. Dis.* **182**, 792–798 (2000).
- McNeil, A. C. *et al.* High-level HIV-1 viremia suppresses viral antigen-specific CD4⁺ T cell proliferation. *Proc. Natl Acad. Sci. USA* **98**, 13878–13883 (2001).
- Pantaleo, G. *et al.* HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease. *Nature* **362**, 355–358 (1993).
- Haase, A. T. *et al.* Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science* **274**, 985–989 (1996).
- Zhang, Z.-Q. *et al.* Kinetics of CD4⁺ T cell repopulation of lymphoid tissues after treatment of HIV-1 infection. *Proc. Natl Acad. Sci. USA* **95**, 1154–1159 (1998).
- Haase, A. T. Population biology of HIV-1 infection: viral and CD4⁺ T cell demographics and dynamics in lymphatic tissues. *Annu. Rev. Immunol.* **17**, 625–656 (1999).
- Geijtenbeek, T. B. *et al.* DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **100**, 587–597 (2000).
- Cheyrier, R. *et al.* Antigenic stimulation by BCG vaccine as an *in vivo* driving force for SIV replication and dissemination. *Nature Med.* **4**, 421–427 (1998).
- Grossman, Z., Feinberg, M. B. & Paul, W. E. Multiple modes of cellular activation and virus transmission in HIV infection: a role for chronically and latently infected cells in sustaining viral replication. *Proc. Natl Acad. Sci. USA* **95**, 6314–6319 (1998).
- Chun, T., Chadwick, K., Margolick, J. & Siliciano, R. Differential susceptibility of naive and memory CD4⁺ T cells to the cytopathic effects of infection with human immunodeficiency virus type 1 strain LAI. *J. Virol.* **71**, 4436–4444 (1997).
- Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl Acad. Sci. USA* **94**, 13193–13197 (1997).
- Fenzi, D. *et al.* Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* **278**, 1295–1300 (1997).
- Carcelain, G. *et al.* Transient mobilization of human immunodeficiency virus (HIV)-specific CD4 T-helper cells fails to control virus rebounds during intermittent antiretroviral therapy in chronic HIV type 1 infection. *J. Virol.* **75**, 234–241 (2001).
- Blancou, P. *et al.* The infiltration kinetics of simian immunodeficiency virus-specific T cells drawn to sites of high antigenic stimulation determines local *in vivo* viral escape. *Proc. Natl Acad. Sci. USA* **98**, 13237–13242 (2001).
- Liszewicz, J. *et al.* Control of HIV despite the discontinuation of antiretroviral therapy. *N. Engl. J. Med.* **340**, 1683–1684 (1999).
- Ortiz, G. M. *et al.* HIV-1-specific immune responses in subjects who temporarily contain virus replication after discontinuation of highly active antiretroviral therapy. *J. Clin. Invest.* **104**, R13–R18 (1999).
- Deeks, S. G. *et al.* Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *N. Engl. J. Med.* **344**, 472–480 (2001).
- Altfield, M. & Walker, B. D. Less is more? STI in acute and chronic HIV-1 infection. *Nature Med.* **7**, 881–884 (2001).
- Dybul, M. *et al.* Short-cycle structured intermittent treatment of chronic HIV infection with highly active antiretroviral therapy: Effects on virologic, immunologic, and toxicity parameters. *Proc. Natl Acad. Sci. USA* **98**, 15161–15166 (2001).
- Waldrop, S. L., Pitcher, C. J., Peterson, D. M., Maino, V. C. & Picker, L. J. Determination of antigen-specific memory/effector CD4⁺ T cell frequencies by flow cytometry: evidence for a novel, antigen-specific homeostatic mechanism in HIV-associated immunodeficiency. *J. Clin. Invest.* **99**, 1739–1750 (1997).
- Hanon, E. *et al.* Fratricide among CD8(+) T lymphocytes naturally infected with human T cell lymphotropic virus type 1. *Immunity* **13**, 657–664 (2000).

Acknowledgements

We thank the patients and staff at UTSW Medical Center and the NIH for their cooperation, and M. Roederer and J. Mascola for their help. This work was supported by the UK Medical Research Council (D.A.P.), the Wellcome Trust, and the National Institutes of Health (S.W.).

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C.D. (e-mail: ddouek@mail.nih.gov).

Reduction of cytochrome *c* oxidase by a second electron leads to proton translocation

Maarten Ruitenberg*, Aimo Kannt†, Ernst Bamberg*, Klaus Fendler* & Hartmut Michel†

* Department of Biophysical Chemistry, Max-Planck-Institut für Biophysik, Kennedyallee 70, D-60596 Frankfurt/Main, Germany

† Department of Molecular Membrane Biology, Max-Planck-Institut für Biophysik, Heinrich-Hoffmann-Strasse 7, D-60528 Frankfurt/Main, Germany

Cytochrome *c* oxidase, the terminal enzyme of cellular respiration in mitochondria and many bacteria, reduces O₂ to water. This four-electron reduction process is coupled to translocation (pumping) of four protons across the mitochondrial or bacterial membrane¹; however, proton pumping is poorly understood. Proton pumping was thought to be linked exclusively to the oxidative phase, that is, to the transfer of the third and fourth electron². Upon re-evaluation of these data, however, this proposal has been questioned^{3,4}, and a transport mechanism including proton pumping in the reductive phase—that is, during the transfer of the first two electrons—was suggested. Subsequently, additional studies reported that proton pumping during the reductive phase can occur, but only when it is immediately preceded by an oxidative phase⁵. To help clarify the issue we have measured the generation of the electric potential across the membrane, starting from a defined one-electron reduced state. Here we show that a second electron transfer into the enzyme leads to charge translocation corresponding to pumping of one proton without necessity for a preceding turnover.

Cytochrome *c* oxidase (see refs 6, 7 for reviews) translocates a total of eight elementary charges per turnover across the membrane; four by pumping four protons across the membrane, and four by taking up the electrons and protons required for the formation of two water molecules from the opposite sides of the membrane. During the catalytic cycle (see Fig. 1a), input of one electron into the oxidized enzyme (state O) leads to the formation of a one-electron reduced form of the enzyme (state E). Upon reduction by a second electron, the enzyme forms state R that can react with oxygen to form state P_M. The two subsequent electron transfers into the cytochrome *c* oxidase convert the enzyme into state F and then back into state O. The putative pathways for electrons and protons in the cytochrome *c* oxidase are shown in Fig. 1b. Electrons are first transferred from cytochrome *c* to the binuclear Cu_A centre at the outer surface, then to haem *a*, and finally to the binuclear haem a₃-Cu_B centre, where oxygen binds and is reduced. The shorter lysine (K)-pathway of proton transfer leads to the active site through K354, whereas the aspartic acid (D)-pathway leads to the active site from D124 via E278. It was thought that only the P_M to F and the F to O transitions are linked to proton pumping², but recent evidence has altered this interpretation from an in-time linkage to an energetic linkage by means of a postulated metastable, energy-rich oxidized form, O~ (ref. 5). The possibility of proton pumping on reduction of the normal oxidized form of the enzyme—which has been proposed on a re-analysis^{3,4} of the published data in ref. 2—has been refuted⁵. This controversy concerns a central property of a fundamental enzyme. Without the correct information on which electron transfers are coupled to proton pumping, any attempts to elucidate the mechanism of proton pumping and its coupling to electron transfer will fail. To provide this basic knowledge, we have incorporated the purified cytochrome *c* oxidase from the soil bacterium *Paracoccus denitrificans* into liposomes, and measured the formation of membrane potentials across the liposomal mem-

brane generated by the injection into the enzyme of single electrons from laser-flash-excited, bound tris(2,2'-bipyridyl)ruthenium^{8,9}.

Generation of a membrane potential by cytochrome *c* oxidase results mainly from three events: (1) electron transfer to haem *a* in the membrane (see Fig. 1b); (2) uptake of a charge-compensating proton¹⁰ from the opposite side of the membrane; and (3) pumping of proton(s) across the entire membrane. Starting from state O, generation of a membrane potential with a time constant of approximately 20 μs has been observed, followed by a deuterium-sensitive phase lasting about 200 μs⁹. The first phase has been attributed to the transfer of one electron to haem *a* (see Fig. 1b). The second phase has been assigned to the uptake of the charge-compensating proton¹⁰ through the K-pathway of proton transfer^{11–13}. No evidence for proton pumping has been observed to date; however, on successive laser flashes, a third phase with a time constant of 1.1 ± 0.3 ms appears, which displays a deuterium-isotope effect⁹ indicating participation of protons. In contrast to our observations, no second phase after the first flash and no third phase upon repetitive flashes were observed in ref. 14. At present we have no explanation for this discrepancy. Unfortunately, a quantitative analysis of the multiple flash experiments is impossible owing

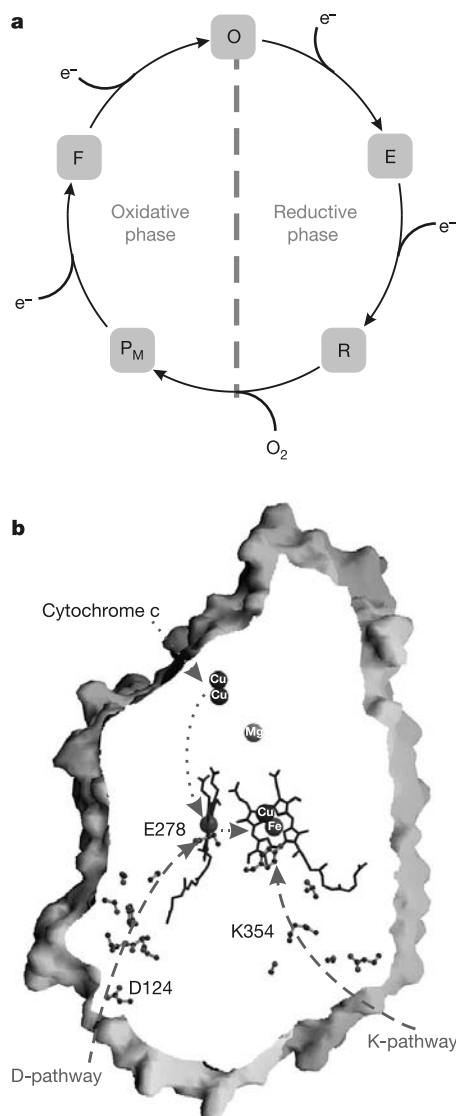


Figure 1 Outline of the catalytic cycle and the structure of cytochrome *c* oxidase. **a**, Simplified version of the catalytic cycle. **b**, Electron and proton transfer pathways in cytochrome *c* oxidase.

to the inhomogeneous state of the enzyme after several flashes⁸. To overcome this problem we have generated a defined, one-electron reduced state (E) by chemical treatment¹⁵. The enzyme in state O is first treated with a (100-fold) molar excess of hydrogen peroxide. The oxoferryl state F ($\text{Fe}^{+IV} = \text{O}^{-II}$) is formed¹⁶. After removal of the excess of hydrogen peroxide the enzyme is treated with carbon monoxide. Carbon monoxide is a two-electron donor and converts state F into a one-electron reduced state E with the haem a_3 -iron reduced according to $\text{Fe}^{+IV} = \text{O}^{-II} + \text{CO} \rightarrow \text{Fe}^{+II} + \text{CO}_2$.

Figure 2 shows the result of this treatment on the spectral properties of the enzyme incorporated into liposomes for the wild type and a solubilized D124N variant. The difference spectrum (E – O) shows maxima at approximately 608 nm, 520 nm and 443 nm, and a minimum at 414 nm, immediately after the treatment with carbon monoxide. Such spectral features have been ascribed to state E (ref. 17). State E is unstable, its maxima disappear and new maxima at 592 nm, 550 nm and 434 nm appear. These new maxima are characteristic for the mixed-valence carbon monoxide (MV-

CO) state¹⁸ in which both haem a_3 and Cu_B of the binuclear centre are reduced and carbon monoxide is bound to the haem a_3 -iron. We find that the speed of the conversion of state E into state MV-CO depends on the enzyme concentration: it is higher at high enzyme concentrations (compare Fig. 2a with b). In addition, state E is much less stable in the solubilized wild-type enzyme compared with the liposome-incorporated one (experiment not shown). These observations may indicate that the enzyme in state E disproportionates according to $2\text{E} \rightarrow \text{O} + \text{R}$. The enzyme in state R then rapidly binds carbon monoxide to form state MV-CO, the enzyme in state O is reduced to state R by carbon monoxide ($\text{O} + \text{CO} + 2\text{H}_2\text{O} \rightarrow \text{R} + \text{CO}_2 + \text{H}_2\text{O}$)¹⁹ which then binds another molecule of carbon monoxide, so that finally the entire enzyme population is converted to state MV-CO. State E is much more stable in the D-pathway variant D124N (see Fig. 2c).

In state E the electron does not reside on a single reduced redox centre but is distributed among haem a , Cu_B and haem a_3 (ref. 17). These redox centres are in a rapid equilibrium²⁰. In the wild-type enzyme, the Soret band absorbance first decreases and finally increases during the $\text{F} \rightarrow \text{E} \rightarrow \text{MV-CO}$ reaction, indicating that in state E the electron must reside to a considerable extent on Cu_B . The D124N variant, however, shows a higher Soret band absorbance in state E, suggesting that compared with the wild-type enzyme, there is a larger fraction of reduced haem. This is supported by the finding of a much higher redox potential (330 compared with 180 mV; P. Hellwig, unpublished data) for haem a in the D124N variant. Thus, in equilibrium there is a larger fraction of reduced haem a , which is also reflected by the increase in intensity and the blueshift of the α -band (Fig. 2). Additional evidence for a fast equilibrium among the redox centres comes from our electrical measurements: although in state E there are differences in electron distribution between the wild-type and variant enzyme, we observe very similar time constants for electron transfer and proton uptake (Table 1, fast and intermediate phase). If electron redistribution between haem a

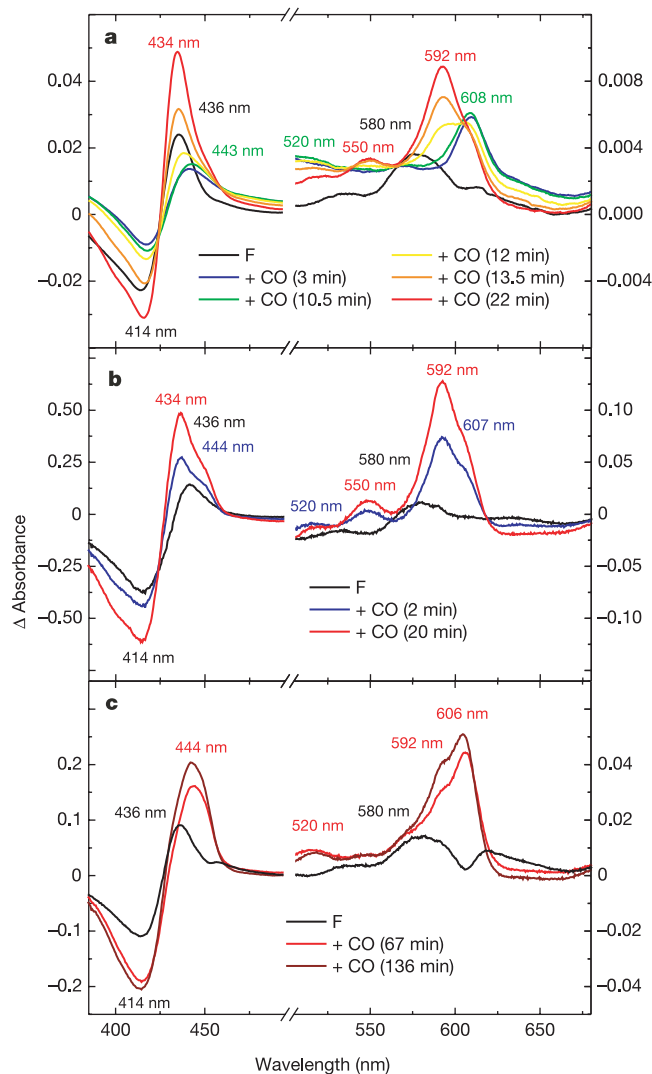


Figure 2 Absorbance difference spectra demonstrating the generation and decay of the one-electron reduced state E. **a**, Wild-type protein concentration at 3 μM . After recording, a reference spectrum (O) hydrogen peroxide and glucose oxidase were added. State F was formed within two minutes. The difference spectrum $\text{F} - \text{O}$ is shown in black. Next, catalase was added and the cuvette was flushed with carbon monoxide in the dark. Difference spectra (minus O) are presented. **b**, As in **a**, but protein concentration was 10 μM . **c**, As in **a**, but solubilized D124N mutant enzyme (5 μM) was used in the presence of detergent.

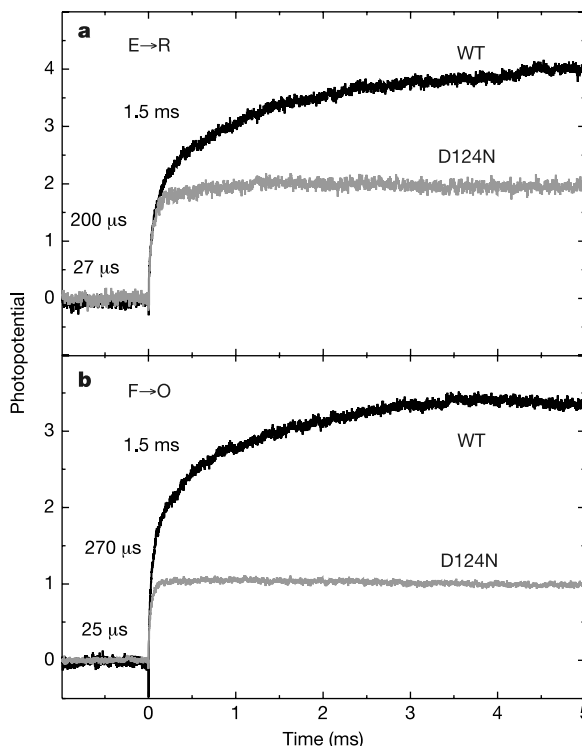


Figure 3 Membrane potential on injection of single electrons into cytochrome c oxidase. **a**, Injection of a single electron into state E of the wild-type (WT) and D124N variant enzyme; **b**, injection of a single electron into state F of the wild-type and D124N variant enzyme. The curves in **a** and **b** were normalized using the amplitudes of the fast phase.

and the binuclear centre was slow, a higher time constant for the fast phase would be expected in the case of the D124N variant. In a tentative model for the reduction of E, the second electron would be rapidly transferred from Cu_A into the haem *a*/haem *a*₃/Cu_A pool, which contains already one electron and which is in rapid equilibrium. At this point the charge-compensating proton is taken up and the electrons might be shifted to the binuclear centre, leading to pumping of an additional proton.

Figure 3a shows the time dependence of the generation of a membrane potential on injecting a single electron into state E for the wild-type and D124N variant enzymes. For the wild-type enzyme, the generation of the membrane potential can be described by three exponentials with time constants $\tau_1 \approx 27 \mu\text{s}$ (fast phase), $\tau_2 \approx 200 \mu\text{s}$ (intermediate phase) and $\tau_3 \approx 1.5 \text{ ms}$ (slow phase), and with relative amplitudes $\alpha_1 \approx 0.26$, $\alpha_2 \approx 0.25$ and $\alpha_3 \approx 0.49$. This behaviour is very similar to the F to O transition (see Fig. 3b). On injecting a single electron into state F, there are also three phases ($\tau_1 \approx 25 \mu\text{s}$, $\tau_2 \approx 270 \mu\text{s}$, $\tau_3 \approx 1.5 \text{ ms}$, with $\alpha_1 \approx 0.3$, $\alpha_2 \approx 0.23$, $\alpha_3 \approx 0.47$, respectively) observed for the formation of the membrane potential for the wild-type enzyme. However, the mutant enzyme D124N behaves differently. For the E to R transition, the slow phase is missing, whereas for the F to O transition both the intermediate and the slow phase are absent. This difference also excludes the possibility that the potential we see developed during the E to R transition is caused by residual presence of state F. Thus, the assignment of the three phases of membrane potential generation seems to be straightforward. The fast phase corresponds to the electron transfer to haem *a*, the intermediate phase reflects the uptake from the opposite side of one proton to compensate for the negative charge of the electron¹⁰, and the slow phase (whose amplitude is approximately the sum of the fast plus the intermediate phases) can now be identified as pumping of one proton across the membrane. The presence of the intermediate phase on reduction of state E in the D124N mutant enzyme might indicate that the charge-compensating proton is taken up by means of the K-pathway, whereas for proton pumping an intact D-pathway is required. The experiments of Fig. 3 allow a clear-cut assignment of the K-pathway to uptake of the charge-compensating proton (in agreement with refs 9, 21), and disprove earlier suggestions based on indirect evidence³.

That we have observed the E to R transition was confirmed as follows. State R is the result of the one-electron reduction of state E. It rapidly reacts with carbon monoxide to form state MV-CO. Photolysis of the haem *a*₃-iron carbon monoxide bond by a laser flash causes rapid electron backflow from haem *a*₃ to haem *a* with a time constant of $3 \mu\text{s}$ (ref. 20), and further backflow to Cu_A at the surface of the membrane (see Fig. 1). This latter backflow causes formation of a membrane potential of opposite sign; however, upon rebinding of carbon monoxide to haem *a*₃, this membrane potential formation is reversed again²² (see Fig. 4a). Successive electron injections into state E enzyme therefore should lead to an increasing

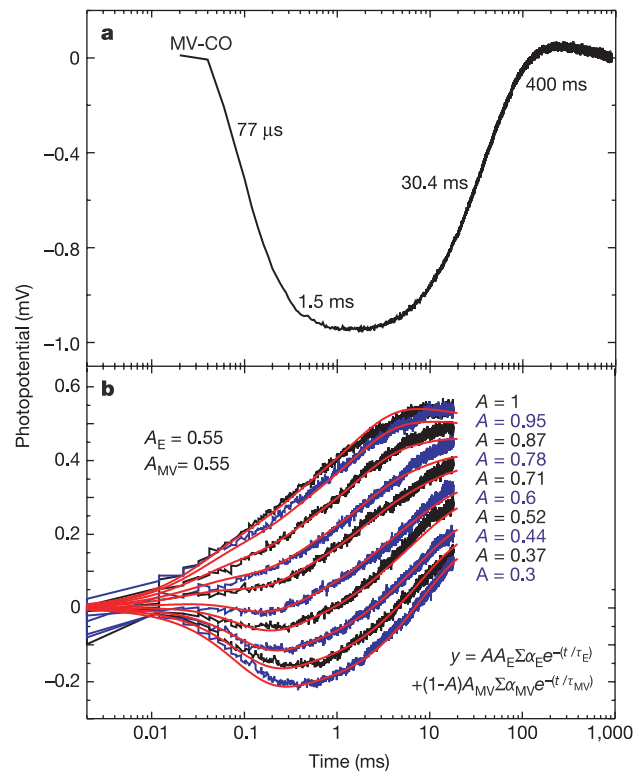


Figure 4 Comparison of photopotentials upon photolysis of state MV-CO and upon successive electron injections into state E. **a**, Photolysis of state MV-CO (presence of tris(2,2'-bipyridyl)ruthenium, pH 9.8, carbon monoxide atmosphere). The response curve can be simulated by four exponentials as presented in Table 1. **b**, Successive electron injections (separated by 1 s) into state E (carbon monoxide atmosphere). The experimental traces (black and blue) can be simulated (red) from the curves in Fig. 3a and Fig. 4a by successively decreasing the percentage of state E from 100% using the kinetic parameters for the curves provided in Table 1. Identity of the values for A_E and A_{MV} is coincidental.

conversion of state E into state MV-CO, and the formation of the membrane potential should change accordingly. Figure 4b shows that this is indeed the case. The time dependence of the generation of the membrane potential can be simulated easily from the parameters (see Table 1) describing the generation of the membrane potential on input of one electron into state E and on laser-flash-induced photolysis of state MV-CO, by varying only the relative contributions of both states.

Our results demonstrate that injection of a single electron into the one-electron reduced state of cytochrome *c* oxidase under anaerobic conditions leads to proton pumping as proposed previously⁴, with similar kinetic constants as the injection of a single electron into the oxoferryl state F. A previous turnover of the

Table 1 Time constants and relative amplitudes for the time dependence of membrane potential generation

Transition	Wild-type enzyme		D124N variant	
	Time constant	Relative amplitude	Time constant	Relative amplitude
E → R	$\tau_1 = 26.9 \pm 7.5 \mu\text{s}$ $\tau_2 = 199 \pm 64 \mu\text{s}$ $\tau_3 = 1.5 \pm 0.6 \text{ ms}$	0.26 ± 0.09 0.25 ± 0.09 0.49 ± 0.09	$\tau_1 = 27.2 \pm 8.9 \mu\text{s}$ $\tau_2 = 195 \pm 13 \mu\text{s}$	0.63 ± 0.10 0.37 ± 0.10
F → O	$\tau_1 = 25.6 \pm 7.7 \mu\text{s}$ $\tau_2 = 269 \pm 105 \mu\text{s}$ $\tau_3 = 1.5 \pm 0.5 \text{ ms}$	0.30 ± 0.08 0.23 ± 0.06 0.47 ± 0.09	$\tau_1 = 24.4 \pm 7.5 \mu\text{s}$	1
MV-CO photolysis	$\tau_1 = 77 \pm 31 \mu\text{s}$ ↓ $\tau_2 = 1.5 \pm 0.7 \text{ ms}$ ↓ $\tau_3 = 30.4 \pm 4.2 \text{ ms}$ ↑ $\tau_4 = 400 \pm 148 \text{ ms}$ ↓	-0.80 ± 0.11 -0.20 ± 0.11 1.20 ± 0.08 -0.20 ± 0.06		

Values for the E to R and F to O transitions, and for the photolysis of the mixed-valence carbon monoxide (MV-CO) enzyme, are shown. The phases of the MV-CO signal have been tentatively assigned to the following processes: τ_1 , electron transfer haem *a* to Cu_A; τ_2 , intracellular H⁺ release; τ_3 , CO-rebinding coupled to electron transfer from Cu_A via haem *a*; τ_4 , discharge of the proteoliposome/planar membrane system.

enzyme is not required for this proton pumping in the reductive phase. Proton pumping only occurs on injection of the second electron. This is notable because in the oxidative phase of the cycle a single electron seems to be sufficient, and may indicate that more redox energy per electron is available in the oxidative part of the cycle. The usage of the proton transfer pathways is partially different. The charge-compensating proton appears to be taken up by means of the K-pathway in the E to R transition (in contrast to the prediction of ref. 4), but through the D-pathway in the F to O transition (Fig. 2b and ref. 21). The D-pathway, however, is required for proton pumping in both cases. The mechanism of proton-pumping itself remains to be elucidated. □

Methods

The proteoliposomes of wild-type or D124N cytochrome *c* oxidase were prepared as described⁹. For the spectroscopic measurements, 500 µl proteoliposomes containing 6 µM or 20 µM enzyme were mixed with 500 µl of 50 mM HEPES/KOH buffer, pH 7.4, 100 mM β-D-glucose in an anaerobic cuvette, degassed and overlaid with argon. After recording a reference spectrum for the oxidized form of the enzyme (O), a 100-fold molar excess of hydrogen peroxide and 40 µg glucose oxidase were added to form state F. Next, we added 25 µg catalase, and the cuvette was flushed with carbon monoxide in the dark. We recorded optical absorbance spectra every 90 s. For comparison the same procedure was used with 10 µM solubilized D124N mutant enzyme in the presence of 0.05% dodecyl-β-D-maltoside as detergent.

The photopotential was measured as described⁹. Proteoliposomes were adsorbed to a planar lipid membrane (protein concentration in the cuvette approximately 100 nM), and the potential was measured across the proteoliposome/planar membrane system. The states E or F were prepared as described above for the spectroscopic measurements. Next, the cytochrome *c* oxidase was reduced upon laser-flash excitation of tris(2,2'-bipyridyl) ruthenium, a photoactivatable electron donor.

Received 30 July 2001; accepted 18 February 2002.

- Wikström, M. K. Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* **266**, 271–273 (1977).
- Wikström, M. Identification of the electron transfers in cytochrome oxidase that are coupled to proton-pumping. *Nature* **338**, 776–778 (1989).
- Michel, H. The mechanism of proton pumping by cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
- Michel, H. Cytochrome *c* oxidase: catalytic cycle and mechanisms of proton pumping—a discussion. *Biochemistry* **38**, 15129–15140 (1999).
- Verkhovsky, M. I., Jasaitis, A., Verkhovskaya, M. L., Morgan, J. E. & Wikström, M. Proton translocation by cytochrome *c* oxidase. *Nature* **400**, 480–483 (1999).
- Ferguson-Miller, S. & Babcock, G. T. Heme/copper terminal oxidases. *Chem. Rev.* **96**, 2889–2907 (1996).
- Michel, H., Behr, J., Harrenga, A. & Kannt, A. Cytochrome *c* oxidase: structure and spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 329–356 (1998).
- Nilsson, T. Photoinduced electron transfer from tris(2,2'-bipyridyl)ruthenium to cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA* **89**, 6497–6501 (1992).
- Ruitenbergh, M. et al. Single-electron reduction of the oxidized state is coupled to proton uptake via the K pathway in *Paracoccus denitrificans* cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA* **97**, 4632–4636 (2000).
- Mitchell, R. & Rich, P. R. Proton uptake by cytochrome *c* oxidase on reduction and on ligand binding. *Biochim. Biophys. Acta* **1186**, 19–26 (1994).
- Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature* **376**, 660–669 (1995).
- Thomas, J. W., Puustinen, A., Alben, J. O., Gennis, R. B. & Wikström, M. Substitution of asparagines for aspartate-135 in subunit I of the cytochrome *bo* ubiquinol oxidase of *Escherichia coli* eliminates proton-pumping activity. *Biochemistry* **32**, 10923–10928 (1993).
- Thomas, J. W., Lemieux, L. J., Alben, J. O. & Gennis, R. B. Site-directed mutagenesis of highly conserved residues in helix VIII of subunit I of the cytochrome *bo* ubiquinol oxidase from *Escherichia coli*: an amphipathic transmembrane helix that may be important in conveying protons to the binuclear center. *Biochemistry* **32**, 11173–11180 (1993).
- Verkhovsky, M. I., Tuukkanen, A., Backgren, C., Puustinen, A. & Wikström, M. Charge translocation coupled to electron injection into oxidized cytochrome *c* oxidase from *Paracoccus denitrificans*. *Biochemistry* **40**, 7077–7083 (2001).
- Witt, S. N. & Chan, S. I. Evidence for a ferryl Fea3 in oxygenated cytochrome *c* oxidase. *J. Biol. Chem.* **262**, 1446–1448 (1987).
- Vygodina, T. V. & Konstantinov, A. A. H₂O₂-Induced conversion of cytochrome *c* oxidase peroxo complex to oxoferryl state. *Ann. NY Acad. Sci.* **550**, 124–138 (1988).
- Moody, A. J. Ligation and electronation states of cytochrome *c* oxidase in relation to other oxidases and peroxidases. *Biochem. Soc. Trans.* **19**, 617–622 (1991).
- Greenwood, C., Wilson, M. T. & Brunori, M. Studies on partially reduced mammalian cytochrome oxidase. Reactions with carbon monoxide and oxygen. *Biochem. J.* **137**, 205–215 (1974).
- Brzezinski, P. & Malmström, B. G. The reduction of cytochrome *c* oxidase by carbon monoxide. *FEBS Lett.* **187**, 111–114 (1985).
- Oliveberg, M. & Malmström, B. G. Internal electron transfer in cytochrome *c* oxidase: evidence for a rapid equilibrium between cytochrome *a* and the binuclear site. *Biochemistry* **30**, 7053–7057 (1991).
- Konstantinov, A. A., Siletsky, S., Mitchell, D., Kaulen, A. & Gennis, R. The roles of two proton input channels in cytochrome *c* oxidase from *Rhodobacter sphaeroides* probed by the effects of site-directed mutations on time-resolved electrogenic intraprotein proton transfer. *Proc. Natl Acad. Sci. USA* **94**, 9085–9090 (1997).

- Jasaitis, A., Verkhovsky, M. I., Morgan, J. E., Verkhovskaya, M. L. & Wikström, M. Assignment and charge translocation stoichiometries of the major electrogenic phases in the reaction of cytochrome *c* oxidase with dioxygen. *Biochemistry* **38**, 2697–2706 (1999).

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie and the Max-Planck-Gesellschaft for financial support. We are grateful to B. Ludwig for providing the D124N mutant.

Correspondence and requests for materials should be addressed to H.M. (e-mail: michel@mpibp-frankfurt.mpg.de) or K.F. (e-mail: fendler@mpibp-frankfurt.mpg.de).

errata

A laboratory analogue of the event horizon using slow light in an atomic medium

Ulf Leonhardt

Nature **415**, 406–409 (2002).

In Table 1 of this Letter, the average particle number for slow light was incorrectly expressed as: $\frac{1}{(e^{\pi\mu} - e^{-\pi\mu})^2}$. It should have read: $\frac{1}{(e^{\pi\mu} + e^{-\pi\mu})^2}$. □

Cbl–CIN85–endophilin complex mediates ligand-induced downregulation of EGF receptors

Phillippe Soubeyran, Katarzyna Kowanetz, Iwona Szymkiewicz, Wallace Y. Langdon & Ivan Dikic

Nature **416**, 183–188 (2002).

In Fig. 3c of this Letter, the line (filled circles) labelled EGF+CIN85 should have been labelled EGFR+CIN85–3SH3, as in Fig. 3d. □

corrigendum

Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine

Tobias Krojer, Marta Garrido-Franco, Robert Huber, Michael Ehrmann & Tim Clausen

Nature **416**, 455–459 (2002).

In this Letter, the Protein Data Bank entry code for the DegP S210A crystal structure is incorrectly listed as 1KJ9. It should be 1KY9. We thank C. Zardecki for bringing this to our attention. □

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Laura Pearson (4977)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silke Opstrup (4994)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau
Tel + 33 (0) 1 43 20 16 51
Fax + 33 (0) 1 43 20 51 52
e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mails: p.phelan@nature.com
k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
e-mail: k.cowan@naturejp.com
Japan Manager: Kate Cowan

naturejobs

Compound disinterest

A dramatic drop in the number of students taking up chemistry in British universities could be seen as both a crisis for the institutions and a career opportunity for those undergraduates who stay the course. A report delivered to the UK government last month highlights a 16% fall in students starting a chemistry degree between 1995 and 2000 — although overall graduate numbers rose by 12% in the same time period.

The figures prompted one former organic chemist, Brian Iddon, now a member of parliament who sits on the Science and Technology Select Committee, to claim that the supply of student chemists has “collapsed” (see *Nature* 416, 777; 2002). But is the same true for demand? Industrial indicators say not. Synthetic organic chemists are still much in demand by pharmaceutical companies, and if the buzz over nanotechnology gets louder, inorganic chemists could see their employment opportunities increase.

The academic indicators are less comforting. Britain's Royal Society of Chemistry has already placed a dozen university chemistry departments on the endangered list. Such a move will not send a positive message to students considering a career in chemistry — especially if they are academically oriented.

So what can be done? For a start, biology professors could mention that students interested in that field may want to consider chemistry as a good starting point — there would be less competition, and strong biochemistry skills are less likely to go out of fashion than molecular-biology techniques. Chemistry professors, meanwhile, could recruit students by following the Massachusetts Institute of Technology's lead and broadcasting the news that plenty of exciting scientific challenges remain in an increasingly unpopulated field.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Norway's boost from
black gold

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Turning oil into science Norway

Norwegian research institutions have historically tried to temper potential international recruits' perceptions about the country's cold, dark winters with talk of its natural beauty and excellent skiing. That approach has not been entirely successful. There are usually more Norwegian scientists leaving than foreign ones arriving. But the new government elected last autumn is planning a different tack.

"I think it's worth to try to attract researchers with the fjords and the mountains," says education and research minister Kristin Clemet. "We need to attract them with good research opportunities and afterwards we can tell them about all the other things."

To create those opportunities, the country is investing the missing ingredient — money. Norway lags far behind its Nordic neighbours Sweden — which spends three times as much per capita on biomedical research — and Denmark, which spends twice as much. At 1.7% of gross domestic product (GDP), Norway's investment in research and development is well below the average of 2.2% given by the Organisation for Economic Co-operation and Development (OECD).

Norway plans to close that gap by tapping into something it has plenty of, at least in the short term — oil. Oil revenues were used to set up the Research Investment Fund in 1999, as a source of extra funding for science. Clemet hopes that the fund, plus incentives for university researchers to commercialize their work

and tax credits to promote industrial investment, will bring Norway's biomedical funding in line with its neighbours, and boost overall R&D up to or above the OECD average.

The fund, at 13 billion kroner (US\$1.7 billion), is ahead of schedule to make its 15-billion kroner target for 2005. The current year is pivotal for science in Norway, as it marks the first time the yield, estimated at 525 million kroner for 2002, will be invested in two large-scale projects: 100 million kroner will be sunk into a functional-genomics initiative and another 100 million will establish 10 centres of excellence.

CENTRE FIELD

Clemet hopes that the two projects will raise the country's international profile and attract foreign researchers — important because the country produces fewer PhDs than other Nordic countries. So, if it wants to grow, it will have to follow Ireland's lead (see *Naturejobs* 4–5; 21 March 2001) and lure Norwegian researchers back as well as tempting scientists from other countries.

The centres, covering four areas of science, face different challenges. Marine research and science that examines the intersection between energy and environment are "obvious" for Norway, says Clemet. The marine biotech business is Norway's second-biggest industry and the country already has a role in sequencing the salmon genome, so it should not be

Kristin Clemet:
emphasizing science over
fjords and mountains.



hard to attract industry interaction and interest from abroad. Environmental science is also targeted for research, as the country already has infrastructure such as the Norwegian Polar Institute.

The other two areas, medicine and health, and information technology, will require more work to build up. Still, they have established a base, says Harald Stenmark, a cell biologist at Radium Hospital in Oslo.

He, like all the scientists in the 40 groups competing to become centres of excellence, is waiting to see if he will make the grade. Centre-of-excellence status, with its average 10 million kroner of funding for up to 10 years, would allow Stenmark to double the number of funded PhD students and postdocs in his lab.

It could also mean that David Gillooly, a postdoc in Stenmark's lab, could stay in the country, which would suit both him and Clemet — Gillooly because he is married to a Norwegian, and Clemet because she has explicitly stated that the centres must attract and retain foreign scientists.

FUGE STATE

In some respects, FUGE, the functional genomics project, which is a bit further along than the centres, is serving as a test case for how well Norway can scale up its science. Clemet likes it because it combines marine, biotech and environmental science. And to succeed, the programme will require both contact with international partners — such as Canada, which is working with Norway to sequence the salmon genome — and industry, which is looking at the science behind many aspects of salmon raising.

“We want to be the leader in the world of salmon genomics,” says Lars Aukrust, a member of the Research Council of Norway. But he thinks that will require more leadership from the government; he and the scientists who proposed FUGE had hoped the project would be funded at three times the rate. “There is no sense of urgency,” Aukrust says of the government.

But Ole Marvik, chairman of the Norwegian Bioindustry Association, says the country does not yet have enough infrastructure to absorb a bigger boost. “You cannot create 500 trained people overnight even if you had the money,” he says.

He notes that the Norwegian biotech industry, with what he calls 45 “true” companies — about half of which are less than three years old — could not sustain excessively rapid growth. Marvik, who is also founder and chief executive of Affitech, an Oslo-based biotech company, has already competed in one bidding war with another biotech company over a recruit and would rather not get into similar battles with universities.

Despite the modest starting size of FUGE and the centres, the two programmes will clearly have an impact on recruitment and employment, says Marvik, with a few hundred positions created immediately, and more commercial spin-offs and academic-industrial collaborations to follow. Although the money is not massive compared with the sums bandied about in the United States, it is enough to encourage more cooperation within Norway and to foster international collaborations. “The amount of money spent is not an important thing alone,” Marvik says. “What’s important is how this investment is creating a momentum.”

Although Werner Christie, chairman of Norway’s Biotechnology Advisory Board, agrees with Aukrust

Low pay, great scenery

In some ways, Fahri Saatcioglu, a researcher at the University of Oslo’s biotech centre, is an example of the kind of person Norway wants more of — he is foreign-born, with ties to universities on both coasts of the United States. But the Turkish scientist’s tenure in Norway also points to the difficulties of attracting more expatriates to the country.

For instance, he says, postdoc stipends are “quite low in Norway, even in comparison to other Nordic countries”.

This makes it difficult to fill posts. Also, there is not a substantial difference between a postdoc’s salary and that of a junior professor — which could deter foreign nationals who are accustomed to a higher pay grade.

“That’s something that really needs to be fixed,” he says. Still, for Saatcioglu, the benefits outweigh the disadvantages. He has learned to ski and enjoys the country’s spectacular scenery. “I moved here with a sense of adventure,” he says.

P.S.

that the rate of investment should be increased, he thinks that the first round of both programmes have given the country enough momentum. “I think we will catch up,” says the former health minister.

But that will require more than just a sustained investment, says Christie, who spent 1999–2001 in San Francisco learning the mechanisms of the bay area’s biotech base. There he witnessed “the power of beer and pizza”, the fuel that drives networking in the United States. “We have to learn that in Norway,” he says.

That educational process has already begun, says Ole Petter Ottersen, science dean of the University of Oslo. Oslo has joined Gothenburg, Sweden, in a collaborative project called Medcoast Scandinavia to form a ‘bioregion’ in the hope of attracting more venture capital — something which the area is short of. He hopes that the nascent bioregion will eventually join forces with the more established Medicon Valley that straddles Copenhagen in Denmark and Lund in Sweden (see *Naturejobs* 10–12; 21 June 2001).

Establishing FUGE and the centres of excellence will play a key role in forming Norway’s part of that broader bioregion. Clemet hopes to speed up both projects if they go well in the first year, perhaps by increasing funding for FUGE and by introducing more centres of excellence. “If we can give researchers good working conditions and interesting challenges, they’ll come even if it’s cold in Norway,” she says.

Paul Smaglik is editor of *Naturejobs*.

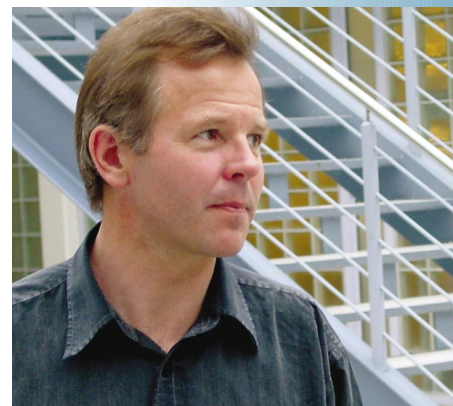
Web links

Norwegian Polar Institute ♦ www.npolar.no

FUGE ♦ www.forskningsradet.no/fag/andre/fuge

Research Council of Norway ♦ www.forskningsradet.no/english

Bioindustry Association ♦ www.biotechforum.no/nba/nba.shtml



Ole Petter Ottersen hopes to establish a new bioregion in Norway.